Supplementary Materials for

Gene transfers, like fossils, can date the tree of life

Material and Methods

Species Trees

All rooted species trees were taken from the literature. The mammalian tree from dos Reis et al[1], the cyanobacterial tree from Szöllősi et al[2]., the archaeal tree from Williams et al[3] . and the fungal tree was taken from Nagy et al[4]. The species trees can be found in
ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/SpeciesTree

Posterior samples for estimating CCPs

Conditional Clade Probabilities[5,67] (CCPs) were estimated from a distribution of gene trees sampled by PhyloBayes. Gene family alignments of the different datasets were taken from the original publications[3,4,8]. For each alignment an MCMC sample was obtained using PhyloBayes (v3.2e)[9] using an LG+Γ4+I substitution model[10] with a burn-in of 1000 samples followed by at least 3000 samples[7,8]. We included all the gene families used in the original publications (Szöllősi et al[2], Williams et al[3] and Nagy et al[4]). The .ale files containing the CCPs for each gene family can be found in
ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/CCPs

Inferring transfers

To infer transfer events we use a hierarchical probabilistic model that models both the evolution of gene phylogenies along the species tree as well as the evolution of sequences along gene trees[11]. We use a joint likelihood that takes into account i) the probability of a given gene tree topology according to a probabilistic model of gene family evolution (a birth-death process that models the duplication, transfer and loss of genes) and ii) the probability of the sequence alignment according to a substitution model. Using such a joint likelihood approach allows one to take into account uncertainty in both gene tree topology and gene tree-species tree reconciliation explicitly during the inference of transfers. Reconciled gene trees, which explicitly imply transfer events used in the downstream reconstruction of relative ages, are sampled according to their joint likelihood using amalgamated likelihood estimation (ALE), a probabilistic approach to exhaustively explore all reconciled gene trees that can be amalgamated as a combination of clades observed in a sample of gene trees. In Szöllősi et al. 2013[7] we demonstrated using simulations that gene trees reconstructed using the above described joint likelihood are substantially more accurate than those reconstructed using sequences alone.
Reconciled gene trees were sampled using the ALE undated[212,7] software (available from the ALE git repository: https://github.com/ssolo/ALE.git)[7]. 100 reconciled gene trees were sampled using ALEml_undated for each family. This allows us to assign a posterior probability to each event (D,T,L) as the fraction of families in which a given event is found. Transfers detected with a posterior probability <0.05 were discarded. Then, for a given transfer between species tree node X and species tree node Y, a global frequency was computed by summing family-wise posterior probabilities across all families. The .uml_recs files containing the reconciled gene trees can be found in
ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/Reconciliations

An example command is:

```
ALEml_undated CyanoSpeciesTree  Family00001.ale 100 _
```

Relative age constraints implied by transfers

The horizontal transfer of genes occurs between contemporaneous species, but this does not mean that the branches inferred as the donor and recipient in reconciled gene trees have to be contemporary[12] (Fig. S8). A transfer leaving the species tree from branch ***a*** and arriving on branch ***b*** will establish a constraint between the father node of branch ***a*** and the daughter node of branch ***b***. This implies that not all transfers carry dating information (Fig. S10). Transfers arriving at leaves or transfers that correspond to constraints implied by the topology (*i.e.* from a node to one of its own descendants) do not carry information on the relative age of nodes in the species tree, and were discarded from the analysis. The data was handled with scripts written in Python 2.7 using extensively the Python library ETE3 [13]

Relative age constraints implied by fossil calibrations

To obtain relative age constraints from fossil calibrations, we used the 26 calibrations from dos Reis et al[1]. These calibrations specify maximum (upper bounds) or minimum age calibrations (lower bounds), or both, for nodes of the species tree. The first step to obtain relative constraints between speciation nodes was propagating lower and upper calibrations up and down the tree respectively. For instance, if a node has no calibration but one of its daughter nodes has a lower calibration then this calibration was propagated up. If a node had an upper calibration, but the daughter node did not, then this was passed down. We then evaluated all pairs of nodes *i,j* that were not in a direct ancestral relationship and recorded an “older” relative constraint if the lower bound calibration of node *i* was older than the upper bound calibration of *j*.

MaxTiC

The MaxTiC[14] algorithm is a heuristic for inferring the relative ages of nodes on a species tree, based on the time orders implied by the largest set of consistent transfers. A set of transfers is called consistent if there exists a time order of the nodes of the species tree with which all transfers from this set are compatible (meaning none of them goes back in time, S9 b). The algorithm takes as inputs a species tree and a set of time-informative transfers (constraints) and outputs a time ranking of speciation nodes that corresponds to the time orders implied by the largest consistent subset of transfers. The sizes of these consistent subsets recovered by MaxTiC in the different datasets are shown Table S2. The python implementation of MaxTiC (available from the ALE github repository at https://github.com/ssolo/ALE.git) was run with the following commands:

```
python maxtic.py species_tree constraint_file ls=180
```

The time orders inferred using MaxTiC were then compared with the ranking of speciation nodes given by the different molecular clock estimates to obtain Figure 3.

Molecular clock estimates of the chronograms

Chronograms were sampled using PhyloBayes 3.3[9] under different clock models (autocorrelated lognormal[9,15], LN; Uncorrelated Gamma Multipliers[16], UGAM; White-noise[17], WN; Strict Clock[9], CL) and two types of prior on the divergence times (Uniform and Birth-death processes)[17]. We used two different models of protein evolution (LG[10] and GTR[18]). An example command is:

```
phylobayes3.3f/exe_lin64/pb -d ./Cyano_alignment.phy -T ./CyanoTree  -x 1 15000
-ugam -unitime -lg -rp 1000 1000 Cyano_UGAM_LG_UNITIME
```

In Archaea we used the alignment of Williams et al[3], which consisted of 10738 amino acid positions, to obtain the species chronogram. We used chronograms sampled from five different chains to obtain the 5000 chronograms used in the Figure 3 (discarding 1000 as burn-in from each chain).[2,7,8]

In the other datasets, where the alignments were much longer, we sampled 4000 columns randomly and used this sampled alignment as an input for PhyloBayes. A single chain was used for each dataset. 5000 chronograms from the beginning of the chain were discarded as burn-in and we sampled every second iteration of the remaining 10000, to obtain a distribution of 5000 chronograms.

In the main figure the model of protein evolution is LG and the prior on divergence times is Uniform. The age of the root was arbitrarily set to 1000. The default options for prior on substitution rates were used. All alignments can be found in ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/Alignments.

Comparison of the clade-to-tip divergence between donors and recipients

To calculate the clad-to-tip divergence between donors and recipients, we fixed the topology of the species tree and then computed branch-lengths in terms of amino acidic substitutions using the concatenates of nearly universal families and a LG+Γ4+I model of protein evolution. These calculations were made using IQ-TREE[19]. Then, we traversed the trees in postorder and computed the divergence of each inner node as the divergence assigned to its children nodes plus the mean branch length of the subtending branches (leaves divergence was set to 0). Then, we took all the constraints found in each data set and separated them into those constraints that had been retained by the MaxTiC algorithm and those that had been discarded. We computed the difference between the estimated clade-to-tip divergence of the donor and recipient clades for each constraint. The trees can be found in: ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/SpeciesTreeSubstitutions

Correlation between fossil-based relative age constraints in mammals and inferred substitutions per site

To calculate the correlation between fossil-based relative age constraints in mammals and inferred substitutions per site, we used the 12 calibration points as given by dos Reis *et al.*[1] that have both lower and upper bounds. We excluded the calibration point at the root, to allow direct comparison with the rankings for transfers (since the ranking of the root is always equal to 0). We also removed the second oldest calibration point because of its high leverage. This resulted in 10 calibration points that were used in the analysis. For each calibration we took the mean point between the upper and lower bounds to obtain the y coordinates. To obtain the substitutions per site (x axis), we took the total branch length separating two species according to the strict molecular clock consensus chronogram and multiplied it by the mean evolutionary rate v (substitution per site / time) estimated by PhyloBayes. The strict molecular clock chronogram was computed on a DNA alignment of universal gene families (used to obtain the species tree in the original paper) without using any calibration points. 15000 chronograms were sampled from a single chain, removing the first 5000 as the burn-in and then resampling one of every two of the remaining 10000 chronograms.

Correlation between the speciation time order and the expected number of substitutions per site in Cyanobacteria

To calculate the correlation between the speciation time order and the expected number of substitutions per site in Cyanobacteria, we measured the speciation ranking given by the MaxTiC algorithm, assigning a value of 1 to the speciation node that is closer to the leaves. The number of substitutions per site was estimated the same way as for the mammalian dataset using the sampled amino-acid alignment described in “Molecular clocks estimates of the chronograms”, the LG model of protein evolution and the rooted species tree topology from Szöllősi *et al.*[8] We sampled 10 speciation nodes, to obtain a correlation comparable to the one previously described in mammals where 10 calibrations had been used. The Spearman’s rank correlation was calculated between the speciation ranking and the expected number of substitutions per site. This procedure was repeated 10000 times. The points, p-value and Spearman’s rho represented in the main figure correspond to the correlation with the median Spearman’s rho of the 10000 correlations. The distribution of p-values can be seen in S13 (lower panel).

Robustness analysis

To assess the support of the transfer-based constraints, we measured the support of each of the relative age constraints using a jackknife approach, sampling without replacement half of the gene families that contain transfers. We repeated this procedure 1000 times and then calculated a set of compatible transfers for each sample. This method also provides us with a measure of the sensitivity of the MaxTiC algorithm to the choice of different gene families. We quantified support for each constraint as the fraction of times that it is observed in the 1000 different random samples. A large fraction (between 40% and 47%) of the constraints appear in at least 95% of the samples (S20-S22 upper panel), indicating that transfer-based constraints are generally robust to variation in the selection of families. We studied how different support thresholds affect the agreement with molecular clock estimates. Varying the threshold we observed a gradual tendency of increasing agreement for higher thresholds (S20-S22, lower panels).

Robustness of inferred relative node ages to an alternative root in Cyanobacteria

Phylogenetic studies using outgroup rooting agree that *Gloeobacter violaceus* is the earliest-diverging lineage within Cyanobacteria.[20–22]9,39,40 This rooting, however, is sensitive to the choice of out-group species[23]. In contrast to these results, recent studies using genome-scale data and alternative rooting methods[8,24] have suggested that *Gloeobacter violaceus* may be a derived lineage within Cyanobacteria. In Figs 3 and 4 we use the root position proposed by Szöllősi et al.10 based on transfers, but we also analyzed the outgroup rooting (see Fig.S24). Comparing the two rootings we found that over 95% (183 out of 191) of the supported relative age constraints under the outgroup root (involving clades found under our alternative root) were also supported under the alternative rooting of Szöllősi et al. In particular, we still recover a recent divergence for the Prochlorococcus-Synechococcus clade. These results, taken together with our simulations[14], suggest that relative age constraints inferred from transfers are robust to uncertainty in the position of the root.

Comparison of information on relative dates from different sources in cyanobacteria

The clade representing heterocyst-forming cyanobacteria (green clade in Fig. S25) is one of the few nodes in our datasets with a clear and ancient fossil calibration. Microfossil evidence from West Africa chert places a minimum bound of 1,957 Mya ("*The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives*" Tomitani et al. PNAS 2006). As shown in the top inset of Fig 1. calibrating only the root of cyanobacteria to be between 3,850 Mya and 2,450 Mya (see e.g. "*Timing of morphological and ecological innovations in the Cyanobacteria: A key to understanding the rise in atmospheric oxygen*" Blank & Sanchez-Baracaldo, Geobiology 2010) produces a severe underestimate of the age of the heterocyst clade. Under a strict molecular clock, including the fossil date as a minimum constraint produces unrealistically narrow confidence intervals and drastically overestimates the age of the unconstrained blue clade, which has been estimated to have diverged 1,020 - 640 Mya based on a dataset with broader taxonomic sampling and additional fossil calibrations (cf. Table 3 in Blank & Sanchez-Baracaldo, Geobiology 2010). This demonstrates that it is necessary to relax the assumption of the strict molecular clock, i.e. to assume that changes in evolutionary rate occur along the phylogeny, with a corresponding increase in model complexity.

The blue and green clade shown in Fig. S25 are particularly relevant because the transfer-based relative age constraints we derive for cyanobacteria in our manuscript constrain the blue clade as well as its three ancestral nodes to be younger than the green clade. Examining the different relaxations of the molecular clock considered in the manuscript we find that the transfer-based constraints are met to different extents by different models (see Table S3 below).

Table S3 demonstrates two important effects: first, as shown in Figure 2a. on the next page using the full agreement score, introducing the internal fossil calibration in the cyanobacterial dataset together with appropriate relaxed clock models can significantly increase agreement with relative transfer-based constraints. This increase in agreement, in fact, provides evidence of congruent dating information in fossil and transfer based constraints in cyanobacteria. Second, however, despite the significant increase in agreement, as can be seen in the last row of Table S3, and more generally in Figure S26a., introducing all available fossil based calibrations and using the best fitting relaxed molecular clock method still does not allow us to sample trees that completely agree with transfer-based constraints.

To determine the limits of agreement between relaxed molecular clocks and transfer-based relative constraints that can be reached under realistic runtimes we launched 10 independent MCMC runs under the best-fitting lognormal model and ran the chains until 30,000 samples were obtained per chain after discarding burn in (approximately one week of computation per chain on 3.7 GHz Intel Xeon processors with PhyloBayes v4.1). Unfortunately, and as shown in Figure S26b., current approaches are limited to sampling chronograms well below 100% agreement. Similar results were obtained for the Fungi and Archaea datasets under extensive sampling (Fungi and Archaea exhibit similar agreement with a median of 0.78 and 0.72 respectively).

Estimating trees calibrated to geological time that carry partial information from transfer-based relative age constraints

To obtain an initial sample of chronograms we used Phylobayes as described above with two additional modifications. First, we added fossil calibrations where available (see Table S4 below). Second, we introduced a subset of the transfer-based relative constraints which involve internally calibrated nodes in the phylogeny by the following procedure: i) for minimum age calibrations all nodes that are constrained to be older by transfer-based relative age constraints were also assigned the same or higher minimum age constraint; ii) conversely, for maximum age constraints that are constrained to be younger by transfer-based relative age constraints we assigned the same or lower maximum age constraint. Introducing this calibration propagation scheme in our datasets we found that it resulted in significantly higher agreement with transfer based constraints in Fungi (median agreement increased from 0.76 to 0.78), but not Cyanobacteria (median 0.729 vs 0.73). For Archaea the lack of internal calibrations prevented us from applying the method. To estimate trees calibrated to geological time that carry partial information from transfer-based relative age constraints we constructed consensus chronograms of the subset of 5% of the sampled chronograms with the highest agreement with transfer-based relative age constraints. Consensus chronograms are shown in Figs. S25-27. The complete set of sampled chronograms and their agreement score can be downloaded from ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/.

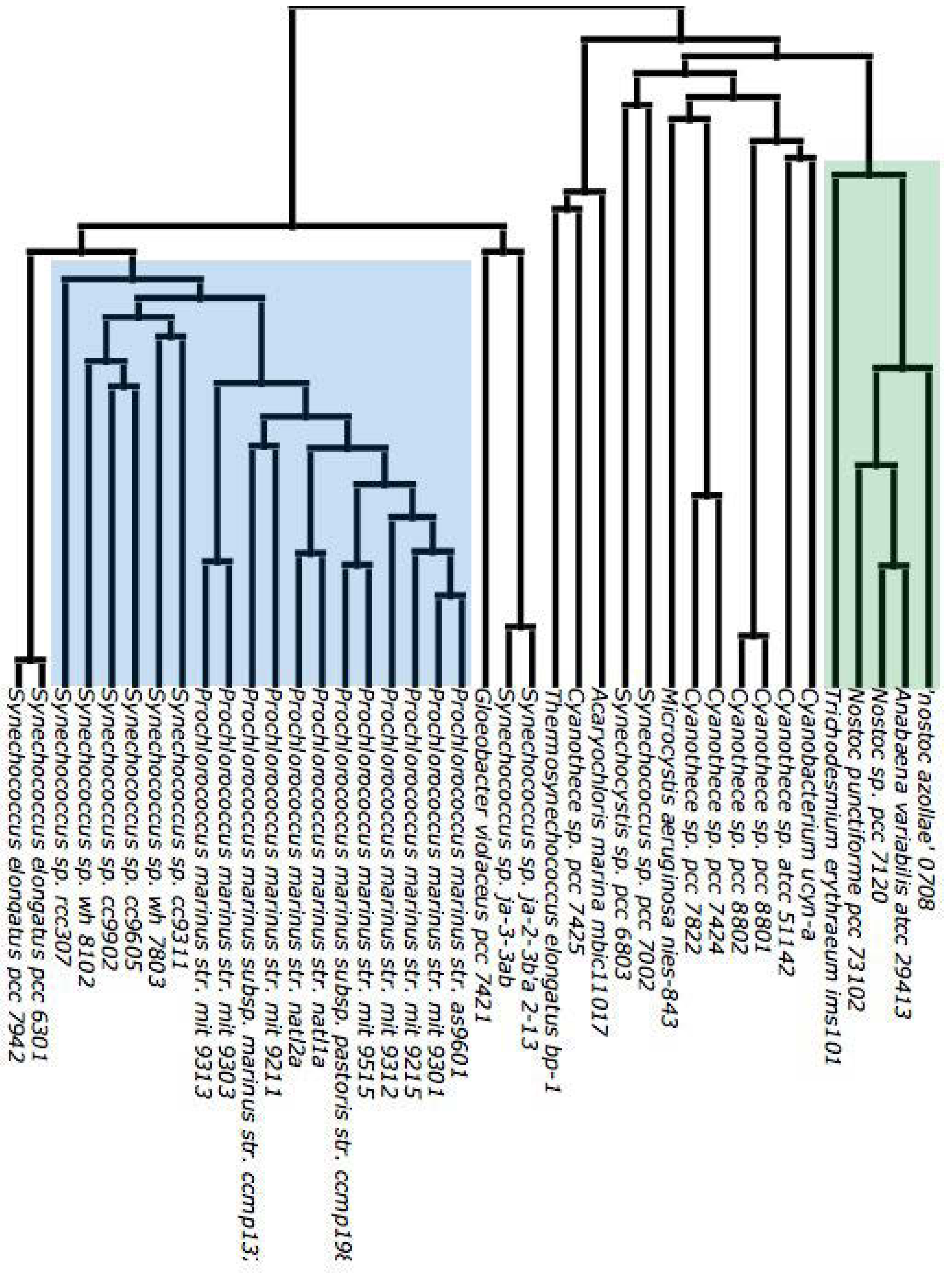
Synechococcus elongatus pcc 7942
Synechococcus elongatus pcc 6301
Synechococcus sp. rcc307
Synechococcus sp. wh 8102
Synechococcus sp. cc9902
Synechococcus sp. cc9605
Synechococcus sp. wh 7803
Synechococcus sp. cc9311
Prochlorococcus marinus str. mit 9313
Prochlorococcus marinus str. mit 9303
Prochlorococcus marinus subsp. marinus str. ccmp13:
Prochlorococcus marinus str. mit 9211
Prochlorococcus marinus str. natl2a
Prochlorococcus marinus str. natl1a
Prochlorococcus marinus subsp. pastoris str. ccmp19(
Prochlorococcus marinus str. mit 9515
Prochlorococcus marinus str. mit 9312
Prochlorococcus marinus str. mit 9215
Prochlorococcus marinus str. mit 9301
Prochlorococcus marinus str. as9601
Gloeobacter violaceus pcc 7421
Synechococcus sp. ja-3-3ab
Synechococcus sp. ja-2-3b'a 2-13
Thermosynechococcus elongatus bp-1
Cyanothece sp. pcc 7425
Acaryochloris marina mbic11017
Synechocystis sp. pcc 6803
Synechococcus sp. pcc 7002
Microcystis aeruginosa nies-843
Cyanothece sp. pcc 7822
Cyanothece sp. pcc 7424
Cyanothece sp. pcc 8802
Cyanothece sp. pcc 8801
Cyanothece sp. atcc 51142
Cyanobacterium ucyn-a
Trichodesmium erythraeum ims101
Nostoc punctiforme pcc 73102
Nostoc sp. pcc 7120
Anabaena variabilis atcc 29413
'nostoc azollae' 0708

Fig. S1: Cyanobacteria species tree. Coloured clades are discussed in the main text. Prochlorococcus - Synechococcus clade in blue and akinete forming cyanobacteria in green.

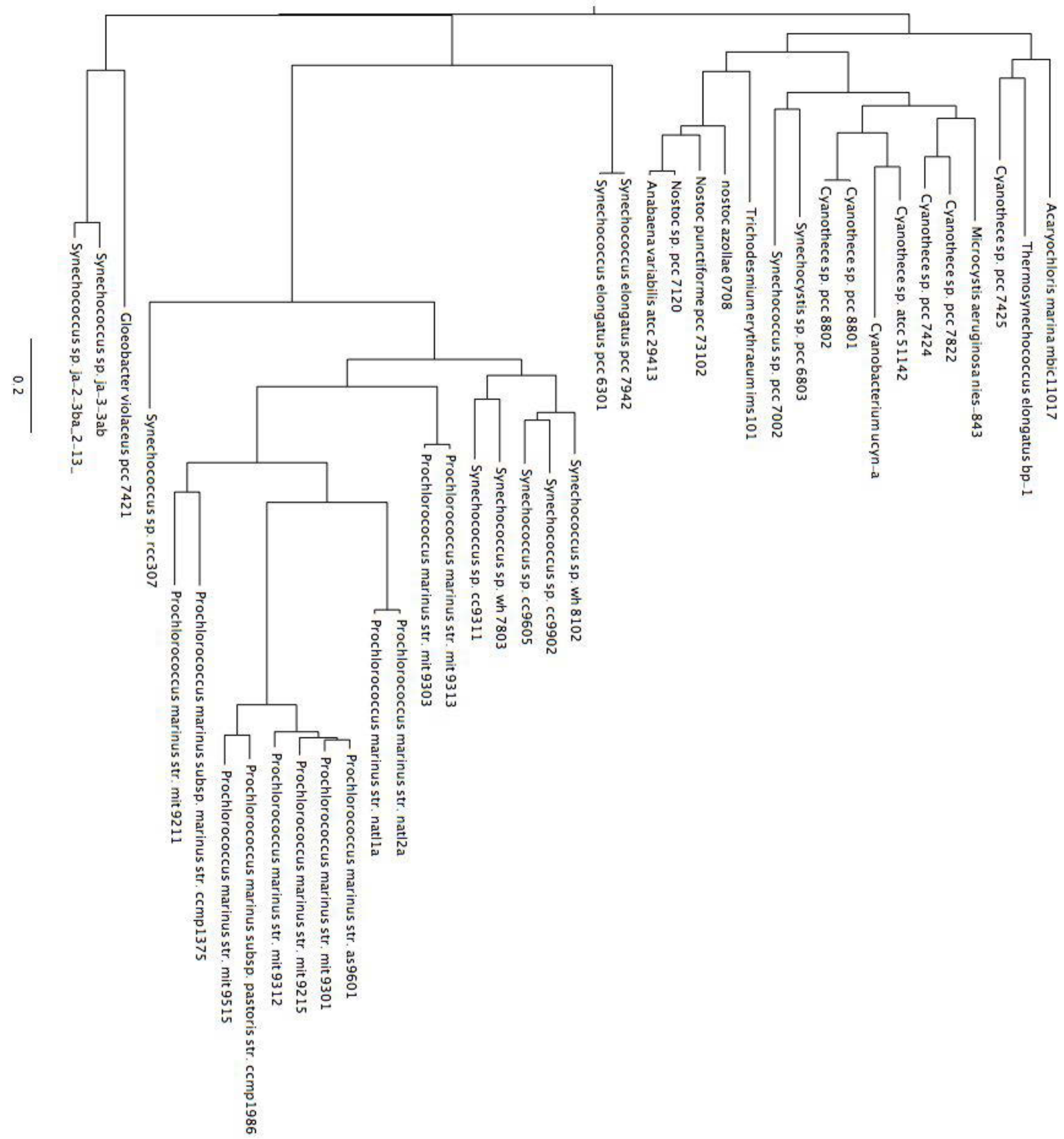
0.2
Synechococcus sp. ja-2-3ba_2-13_
Synechococcus sp. ja-3-3ab
Gloeobacter violaceus pcc 7421
Synechococcus sp. rcc307
Prochlorococcus marinus str. mit 9211
Prochlorococcus marinus subsp. marinus str. ccmp1375
Prochlorococcus marinus str. mit 9515
Prochlorococcus marinus subsp. pastoris str. ccmp1986
Prochlorococcus marinus str. mit 9312
Prochlorococcus marinus str. mit 9215
Prochlorococcus marinus str. mit 9301
Prochlorococcus marinus str. as9601
Prochlorococcus marinus str. natl1a
Prochlorococcus marinus str. natl2a
Prochlorococcus marinus str. mit 9303
Prochlorococcus marinus str. mit 9313
Synechococcus sp. cc9311
Synechococcus sp. wh 7803
Synechococcus sp. cc9605
Synechococcus sp. cc9902
Synechococcus sp. wh 8102
Synechococcus elongatus pcc 6301
Synechococcus elongatus pcc 7942
Anabaena variabilis atcc 29413
Nostoc sp. pcc 7120
Nostoc punctiforme pcc 73102
nostoc azollae 0708
Trichodesmium erythraeum ims101
Synechococcus sp. pcc 7002
Synechocystis sp. pcc 6803
Cyanothece sp. pcc 8802
Cyanothece sp. pcc 8801
Cyanobacterium ucyn-a
Cyanothece sp. atcc 51142
Cyanothece sp. pcc 7424
Cyanothece sp. pcc 7822
Microcystis aeruginosa nies-843
Cyanothece sp. pcc 7425
Thermosynechococcus elongatus bp-1
Acaryochloris marina mbic11017

Fig. S2: Cyanobacteria species tree with branch length proportional to the number of substitutions, computed using IQ-TREE. This tree was used to compute the clade-to-tip divergence of donors and recipients for transfer events.

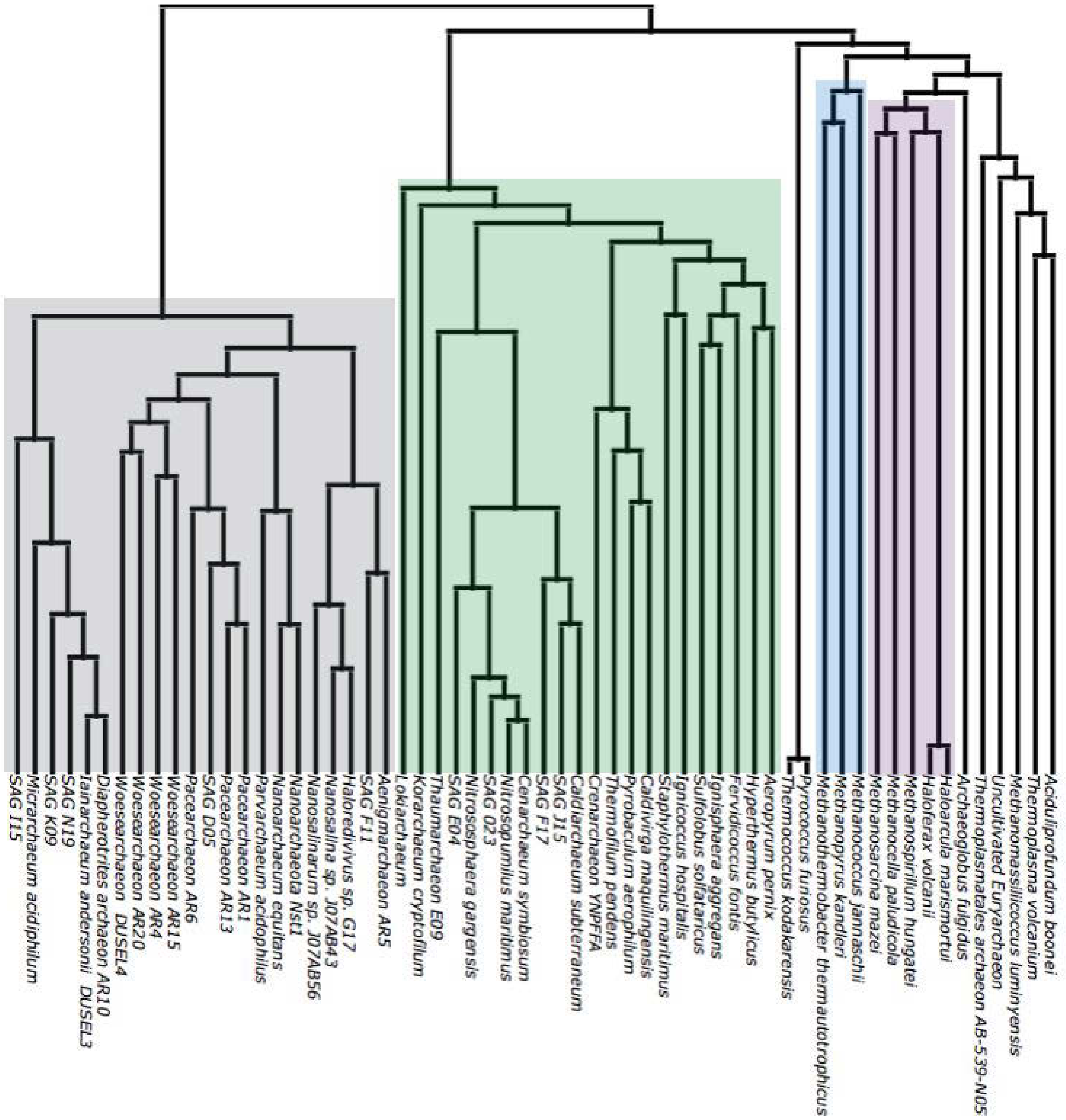

Fig. S3: Archaeal species tree**.** Coloured clades are discussed in the main text. Cluster 1 methanogens in blue, cluster 2 methanogens in purple, TACK+*Lokiarchaeum* in red and and DPANN in grey

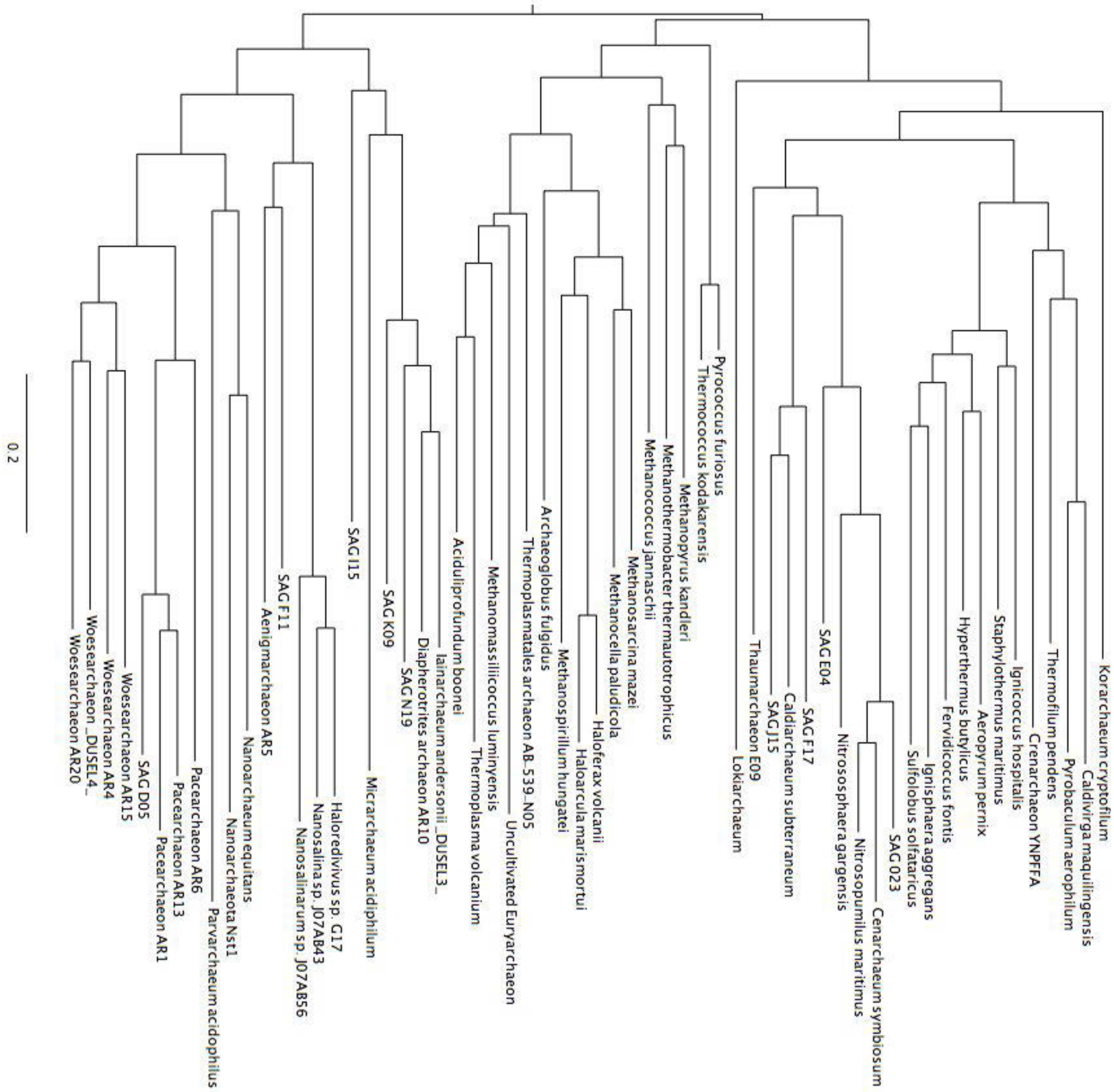

Fig. S4: Archaea species tree with branch length proportional to the number of substitutions, computed using IQ-TREE. This tree was used to compute the clade-to-tip divergence of donors and recipients for transfer events.

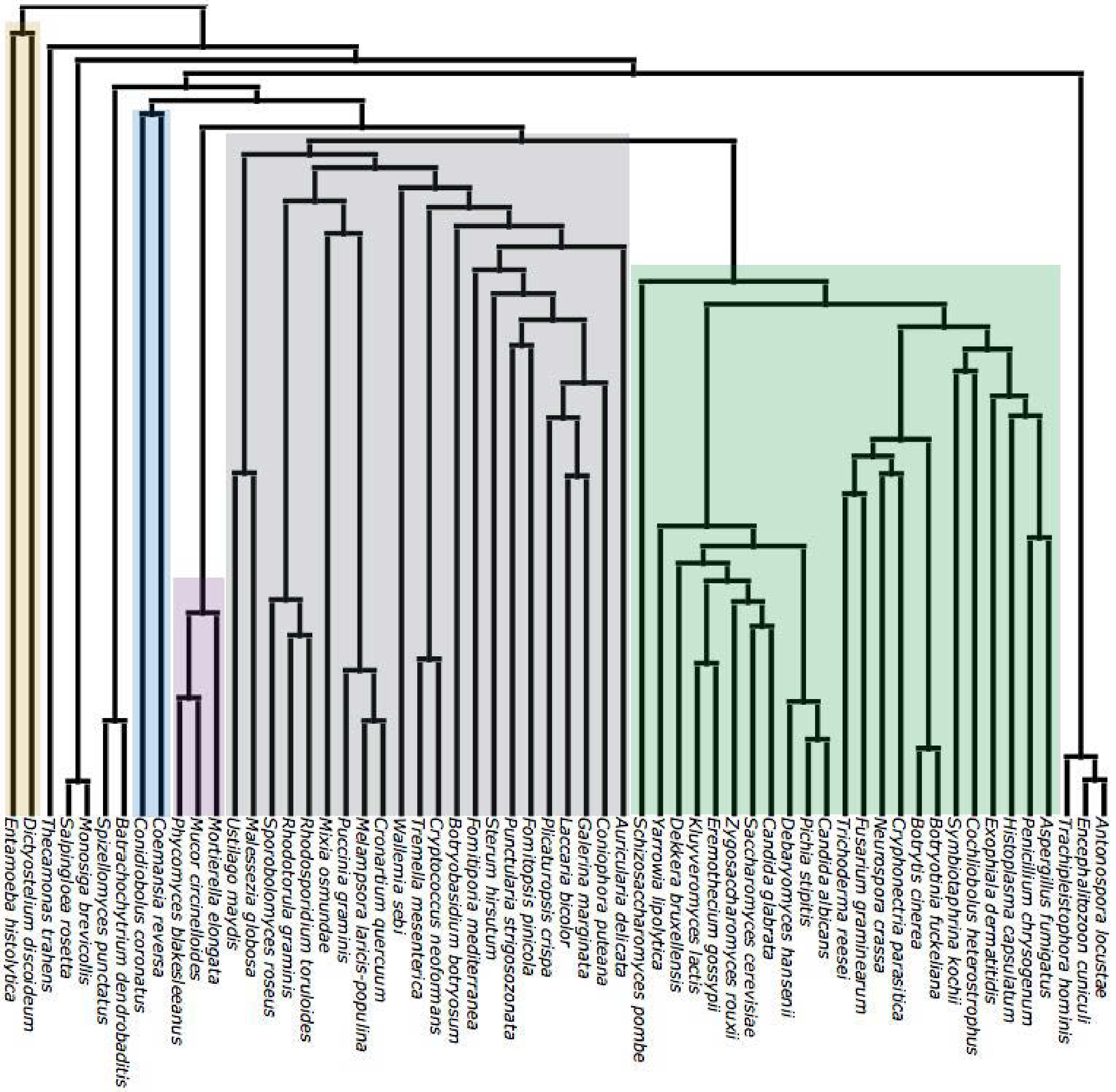

Fig. S5: Species tree of Fungi. Coloured clades are discussed in the main text. Amoebozoa in yellow, Zoopagomycota in blue, Mucuromycotina in purple, Basidiomycota in grey and Ascomycota in green.

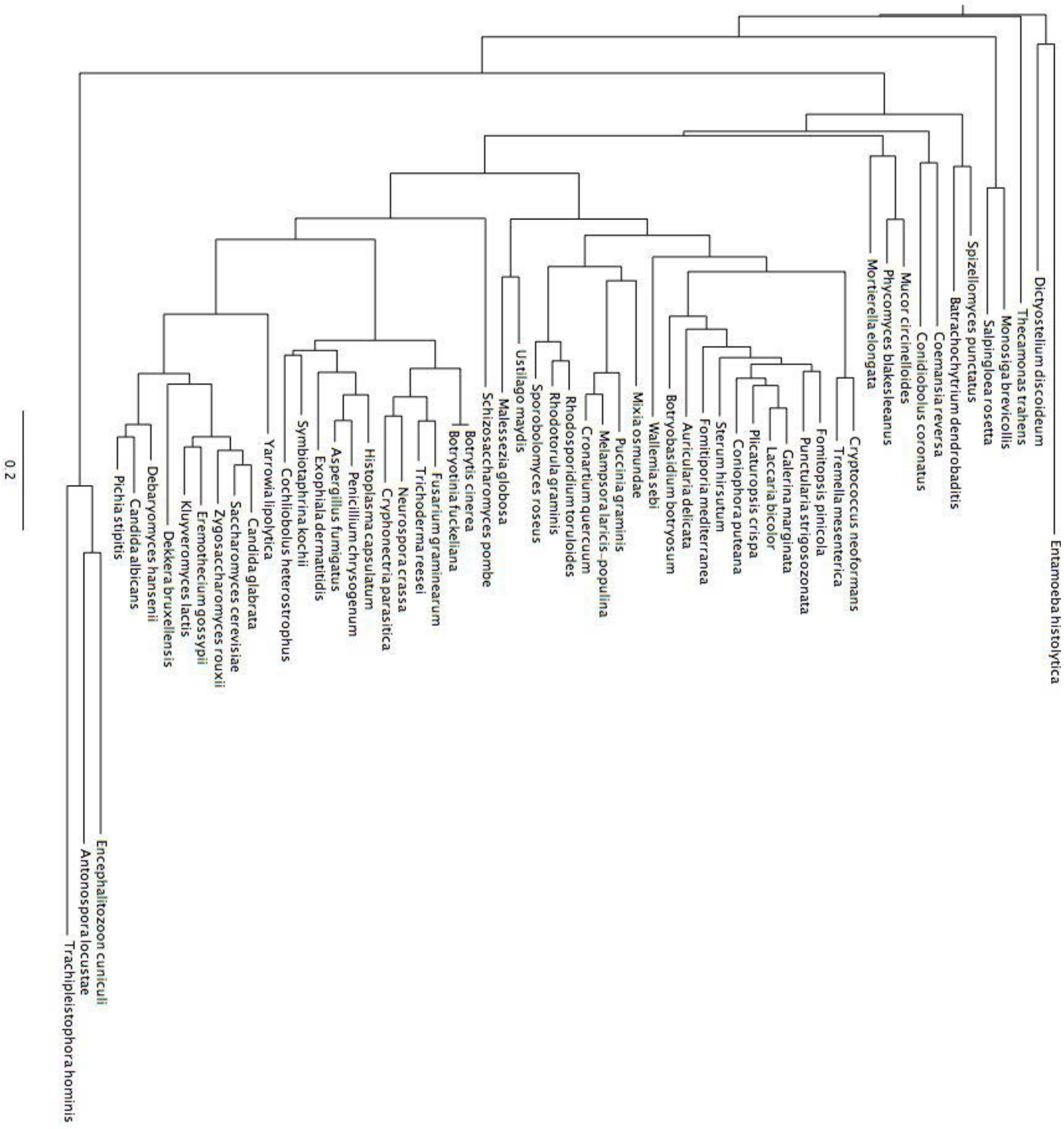

Fig. S6: Fungi species tree with branch length proportional to the number of substitutions, computed using IQ-TREE. This tree was used to compute the clade-to-tip divergence of donors and recipients for transfer events.

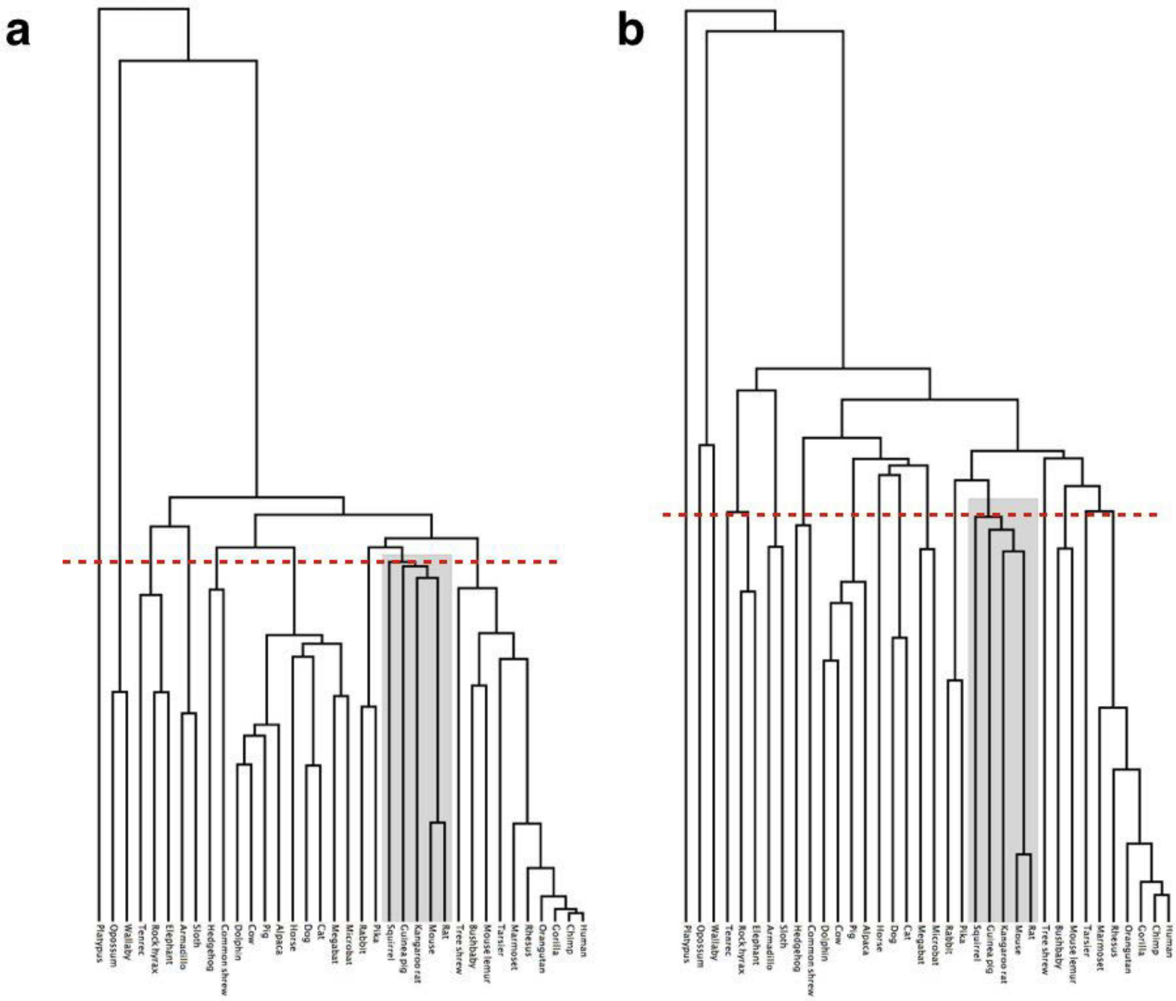

Fig. S7: Mammal species tree with annotated inner nodes (strict molecular clock consensus chronogram in **a**, and lognormal consensus chronogram in **b**). Rodents are highlighted in grey.

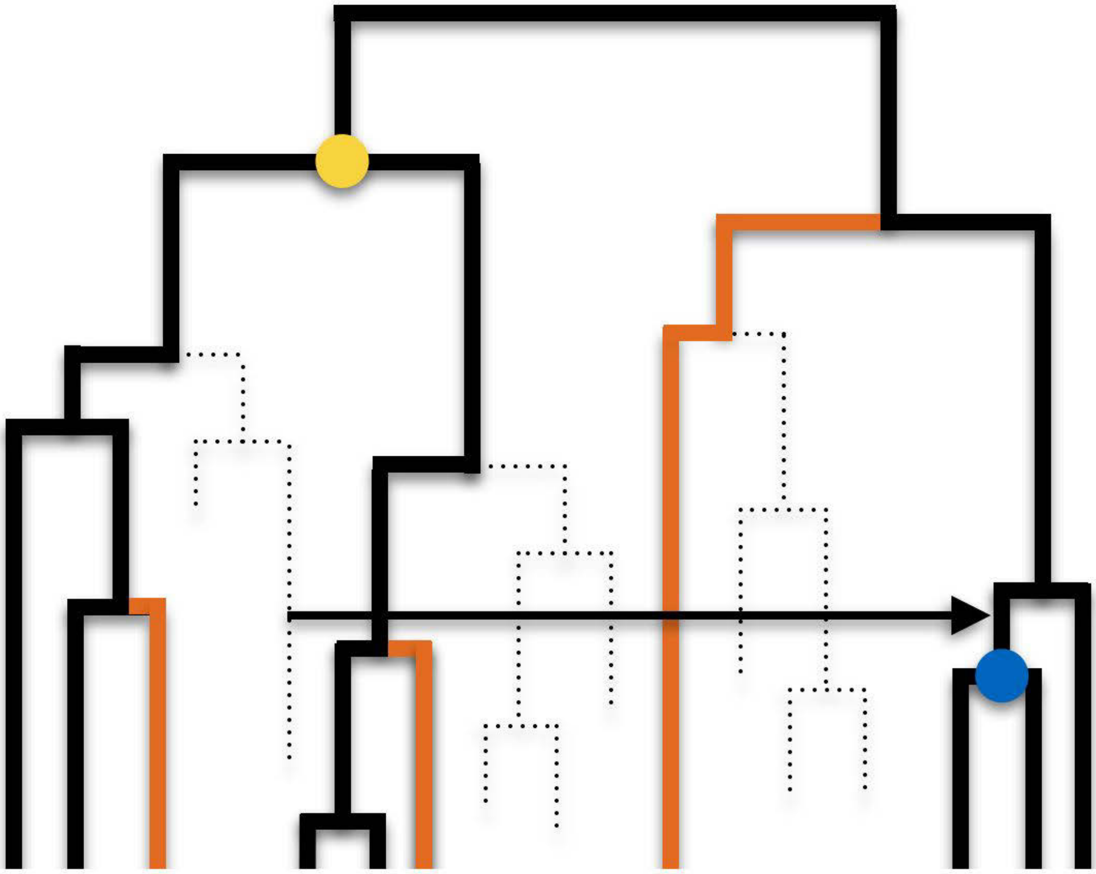

Fig. S8: Transfers occur between contemporaneous species. Thick black lines show the phylogeny of sampled species included in the phylogenetic analysis. Orange lines represent extant but unsampled lineages. Dotted black lines represent extinct species. This figure is illustrative, because in real data the proportion of extinct and unrepresented clades compared to sampled diversity is much higher[12]. An LGT event is indicated by the arrow. The donor lineage for the LGT becomes extinct, but the LGT survives in descendants of the recipient lineage. Note that the apparent donor species lineage on the represented tree is not contemporaneous to the recipient species tree branch where the transfer arrives. This transfer event implies the constraint that the yellow node is older than the blue node.

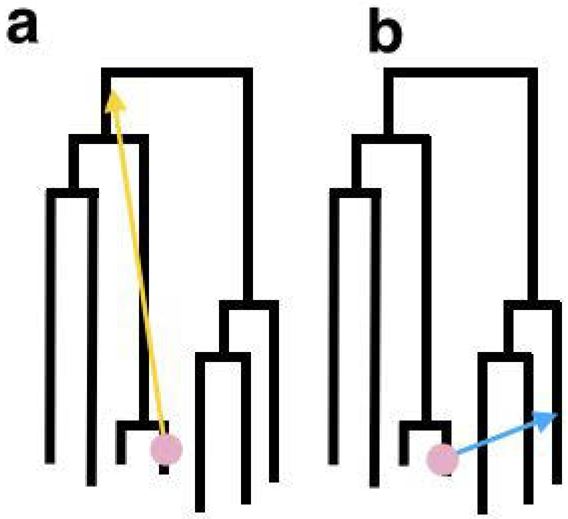

Fig. S9: Reconciliations considered by ALE undated. ALE undated uses a series of gene Duplication, Transfer and Loss events to reconcile a gene tree with the species tree[11]. In contrast to dated DTL methods, where transfers can only occur between branches that are contemporary, in the undated method transfers can occur between any two branches in the tree (e.g. transfer in b is allowed), with the exception of transfer where the recipient branch is an ancestor of the donor branch[2,25] (e.g. transfer in a is forbidden)

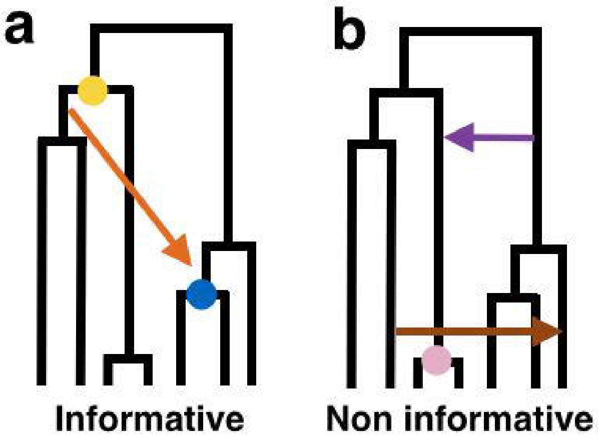

Fig. S10: Informative vs. non-informative transfer. The informative transfer above (panel a, same transfer as in Fig. S8) indicates that the yellow node must be older than the blue node. Non-informative transfers (b) fall in two categories. The non-informative transfer in brown is transferred into a leaf; since leaves have no descendants, such transfers do not imply a relative age constraint. The transfer highlighted in purple is non-informative because it implies a constraint that is already established by the position of the root.

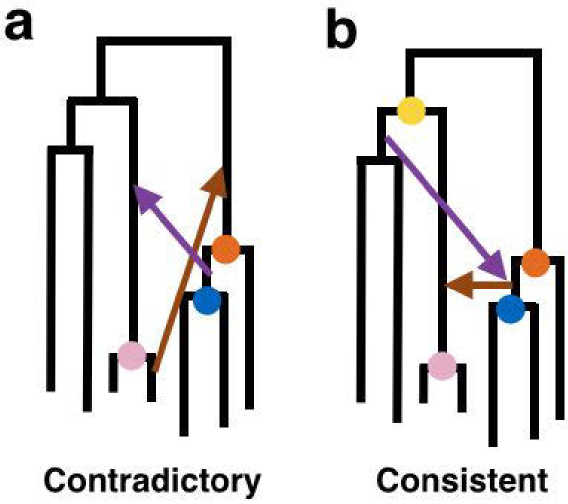

Fig. S11: Contradictory vs. consistent transfers. The transfers in **a** are contradictory because they imply different orders of speciations. The brown transfer indicates that the pink node must be older than the orange node. The purple transfer indicates that the orange node must be older than the green node. Transfers in **b**, on the other hand, are consistent because there exists an ordering of speciations that is compatible with the constraints implied by the transfers. The purple transfer implies that the yellow node must be older than the blue node and the brown transfer implies that the orange node must be older than the pink node.

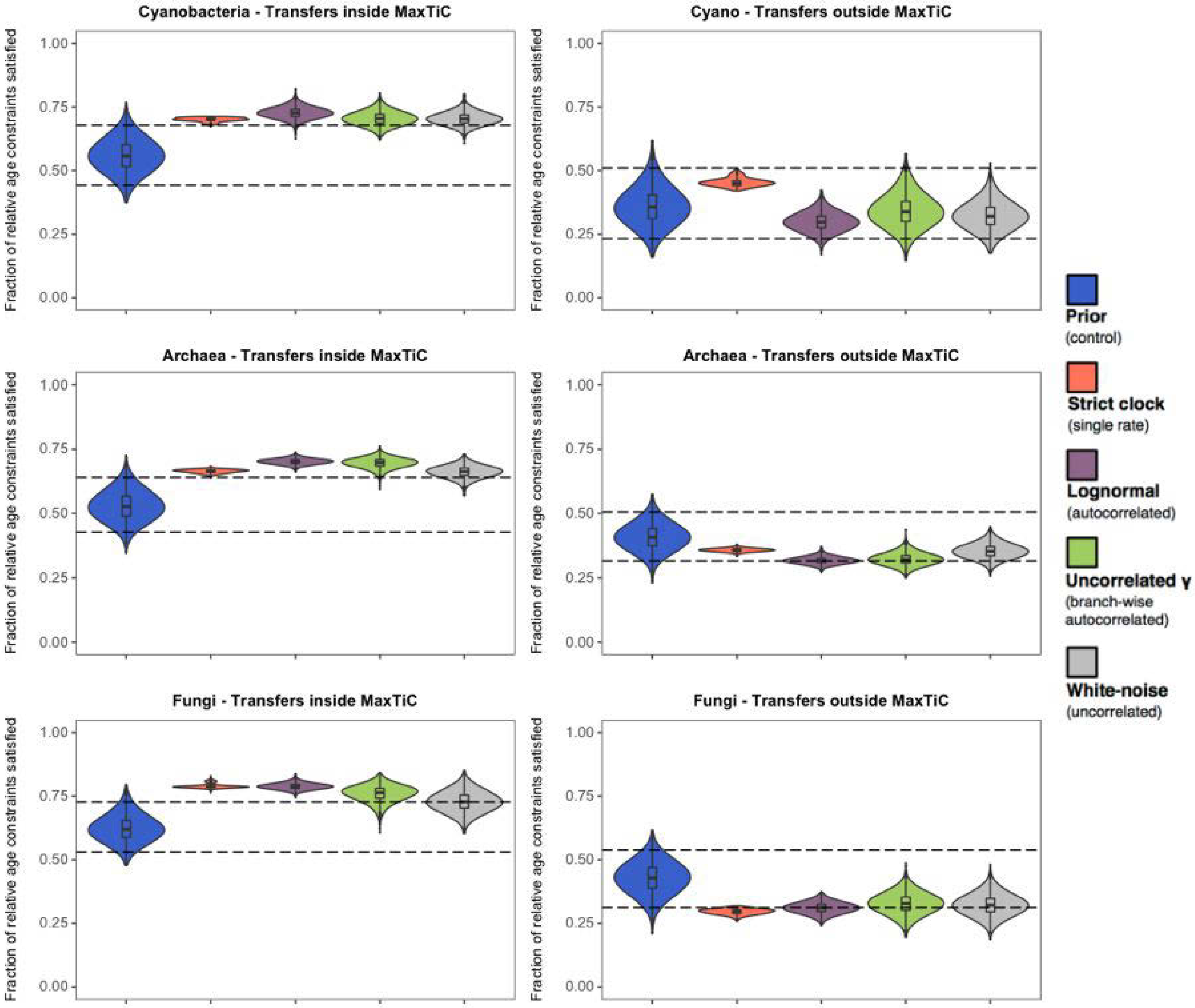

Fig. S12: Agreement between Molecular Clock estimates and the constraints established by transfers selected (left column) and discarded (right column) by the MaxTiC algorithm. The left column shows the same data as Figure 2, and is used here as a comparison on the same scale as the right column. The agreement score of the prior distribution (blue) changes between the two sets of constraints because on the left the set of selected transfers is consistent by construction (*i.e.* agreeing with at least one chronogram), while on the right the set of discarded transfers is not necessarily consistent. Taking this into account, it appears that molecular clocks tend to agree with the transfers selected by the MaxTiC (left) and disagree, or at least not significantly agree, with the discarded transfers (right).

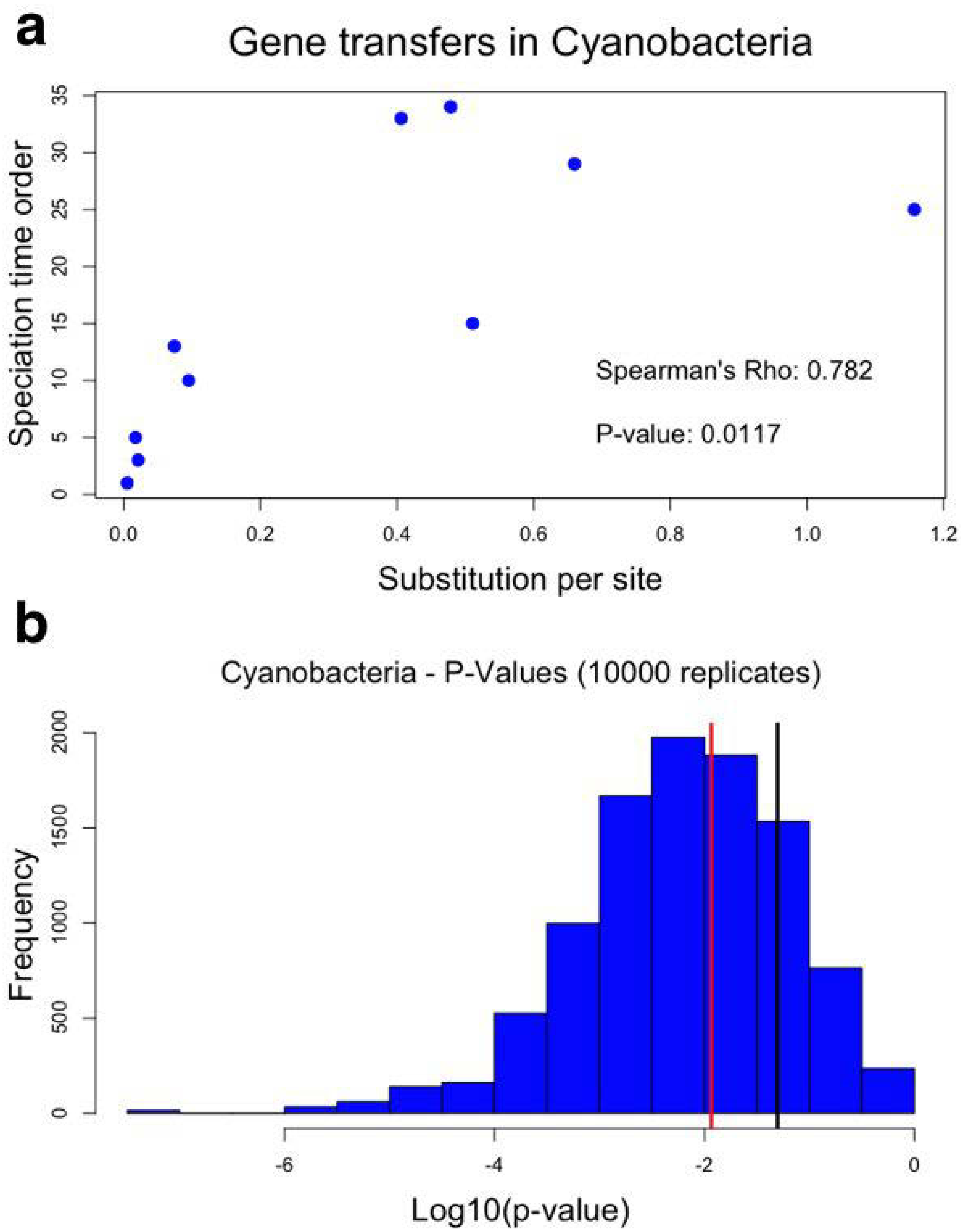

Fig. S13: Correlation between speciation time order indicated by transfers and substitutions per site (Cyanobacteria). The p-value of the correlation in **a** is indicated by a red line in **b**. The significance level at 0.05 is indicated by a black vertical line in **b**

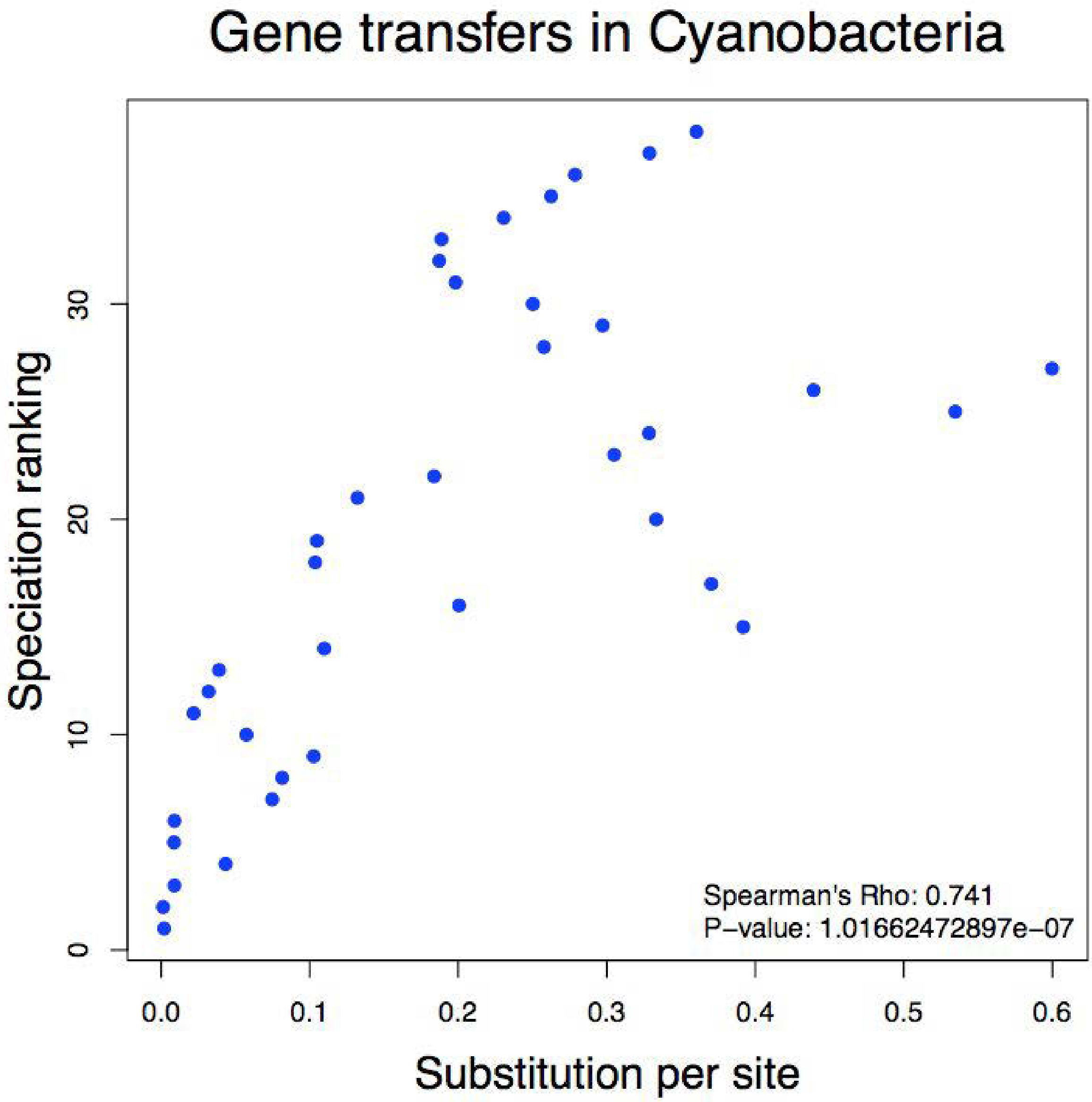

Fig. S14: Correlation between speciation time order indicated by transfers and substitutions per site for all speciation nodes (Cyanobacteria).

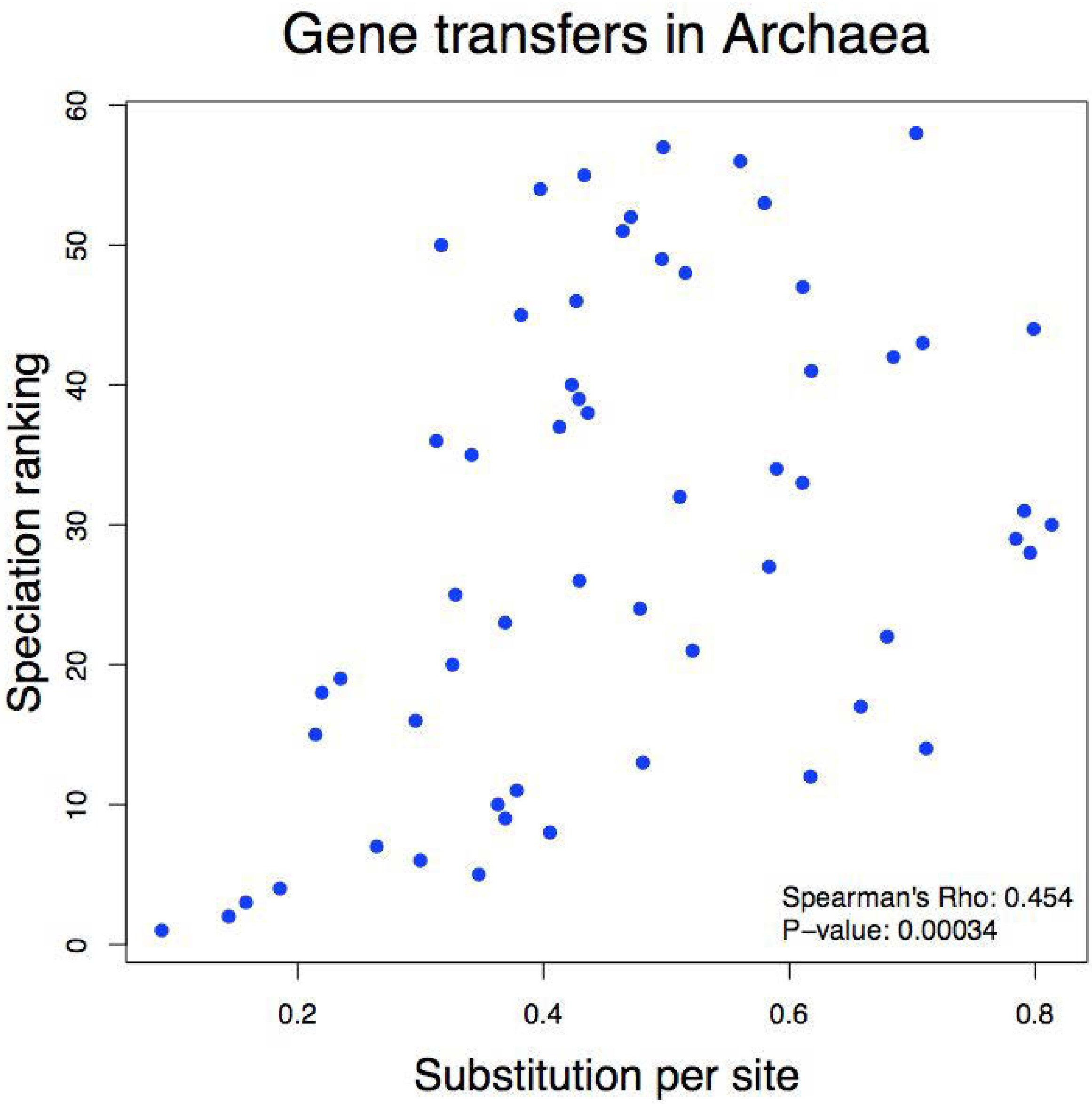

Fig. S15: Correlation between speciation time order indicated by transfers and substitutions per site for all speciation nodes (Archaea).

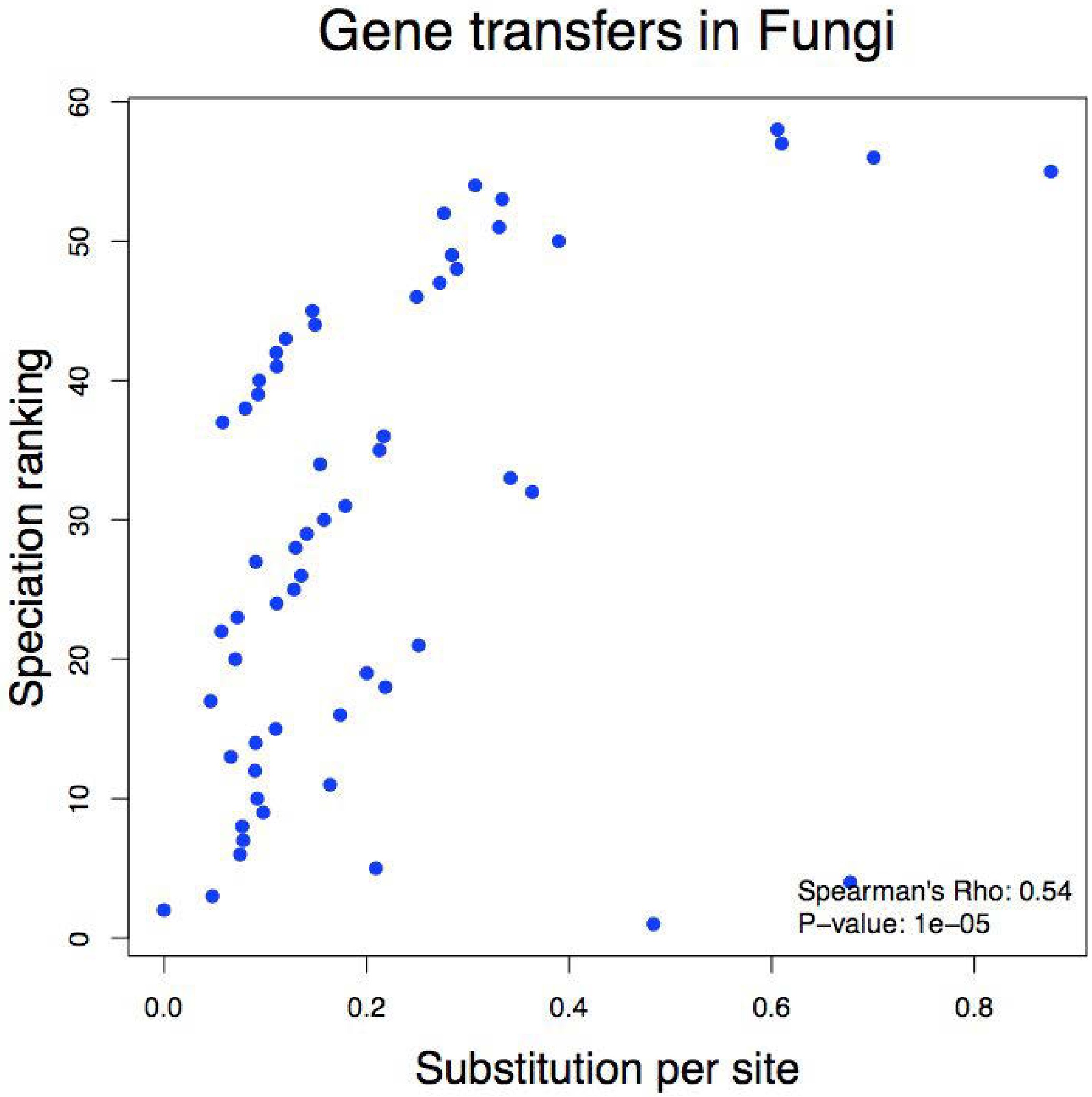

Fig. S16: Correlation between speciation time order indicated by transfers and substitutions per site for all speciation nodes (Fungi).

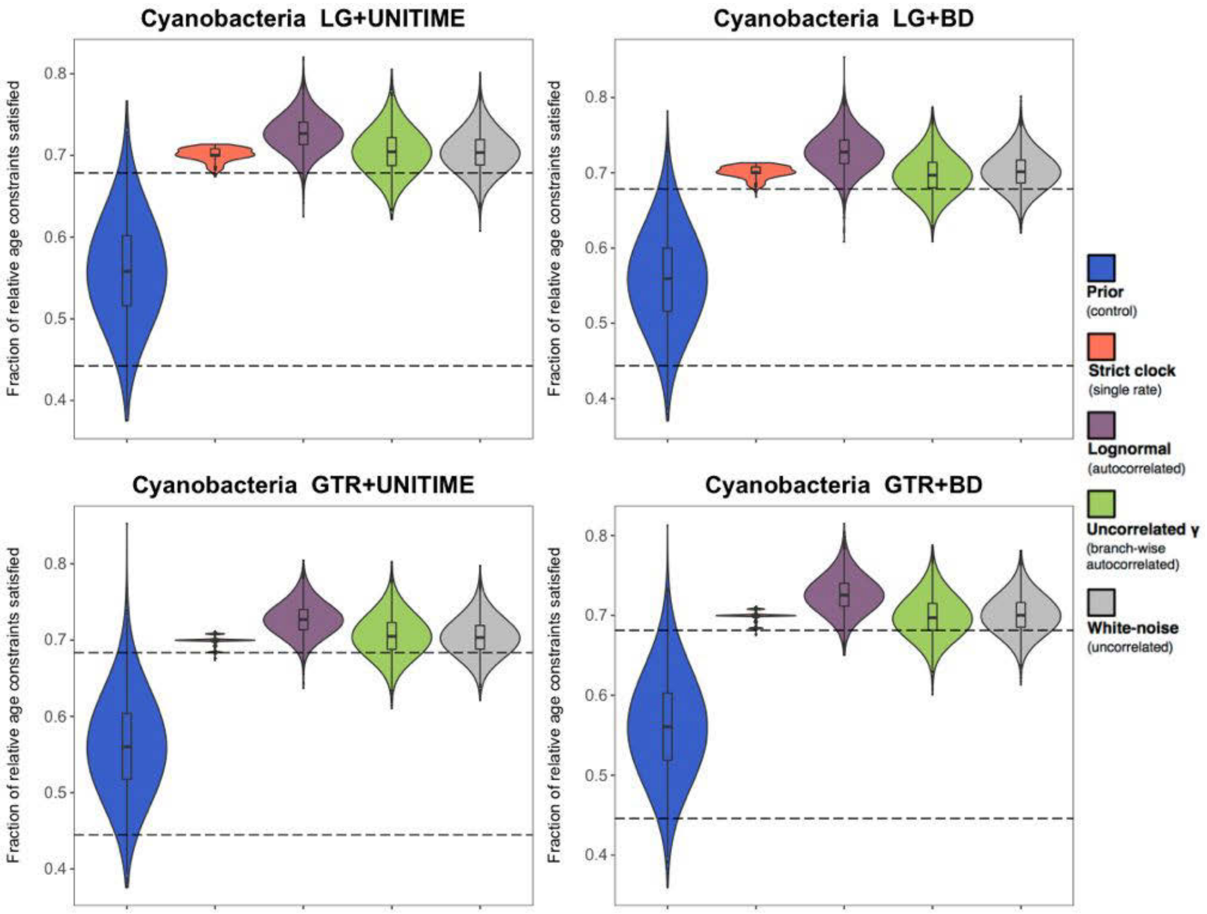

Fig. S17: Agreement between clocks and transfer-based constraints in Cyanobacteria. Different substitution models (LG and GTR) were used as well as different priors (uniform ~ UNITIME and Birth-Death ~ BD). The most noticeable effect seems to be a slightly different behaviour of the strict clock when using different models of protein evolution, but the results remain qualitatively similar. Colors correspond to the different types of clocks, as in Fig. 3 (orange strict clock, purple lognormal, green uncorrelated gamma-multipliers and gray white-noise). The area delimited by the dashed lines corresponds to the 95% confidence interval determined by the prior.

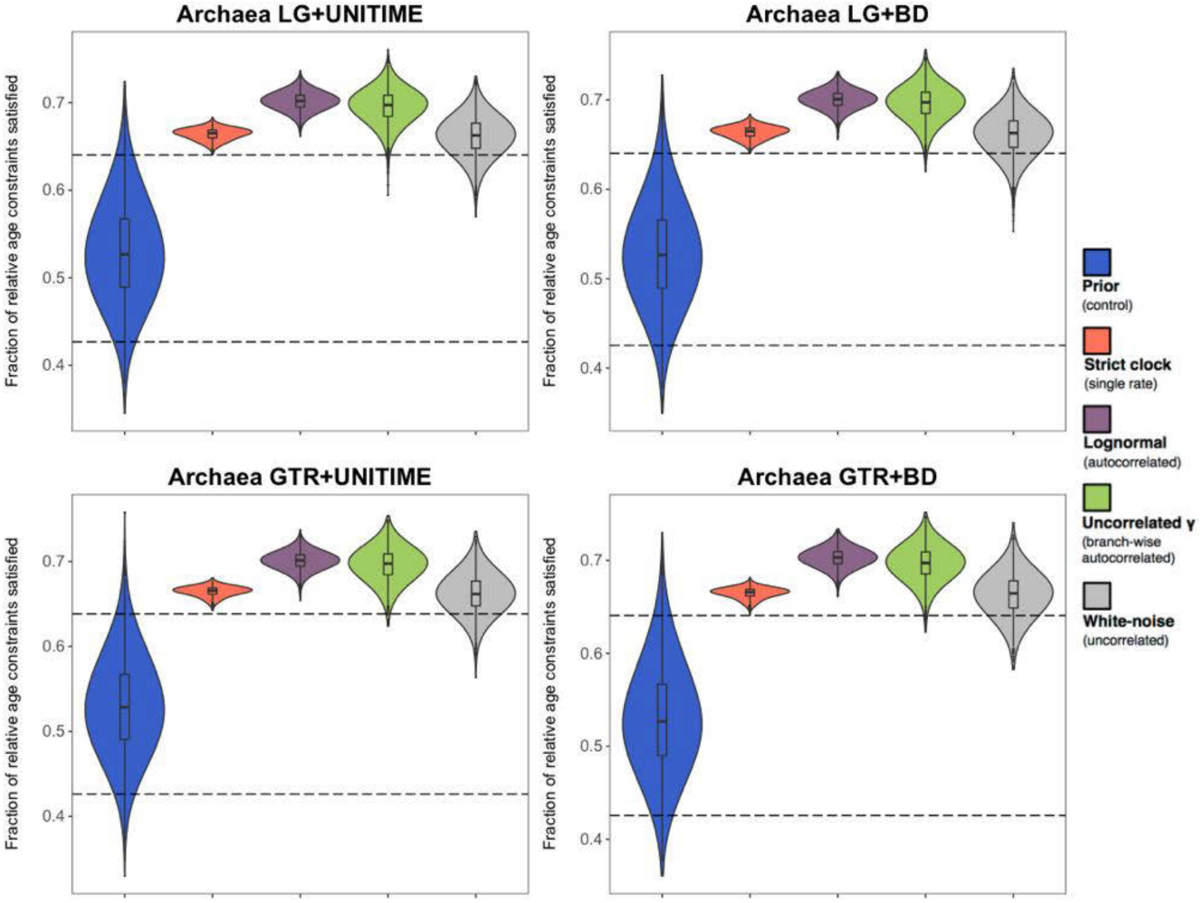

Fig. S18: Agreement between clocks and transfers based-constraints for Archaea. See text in Fig. S17.

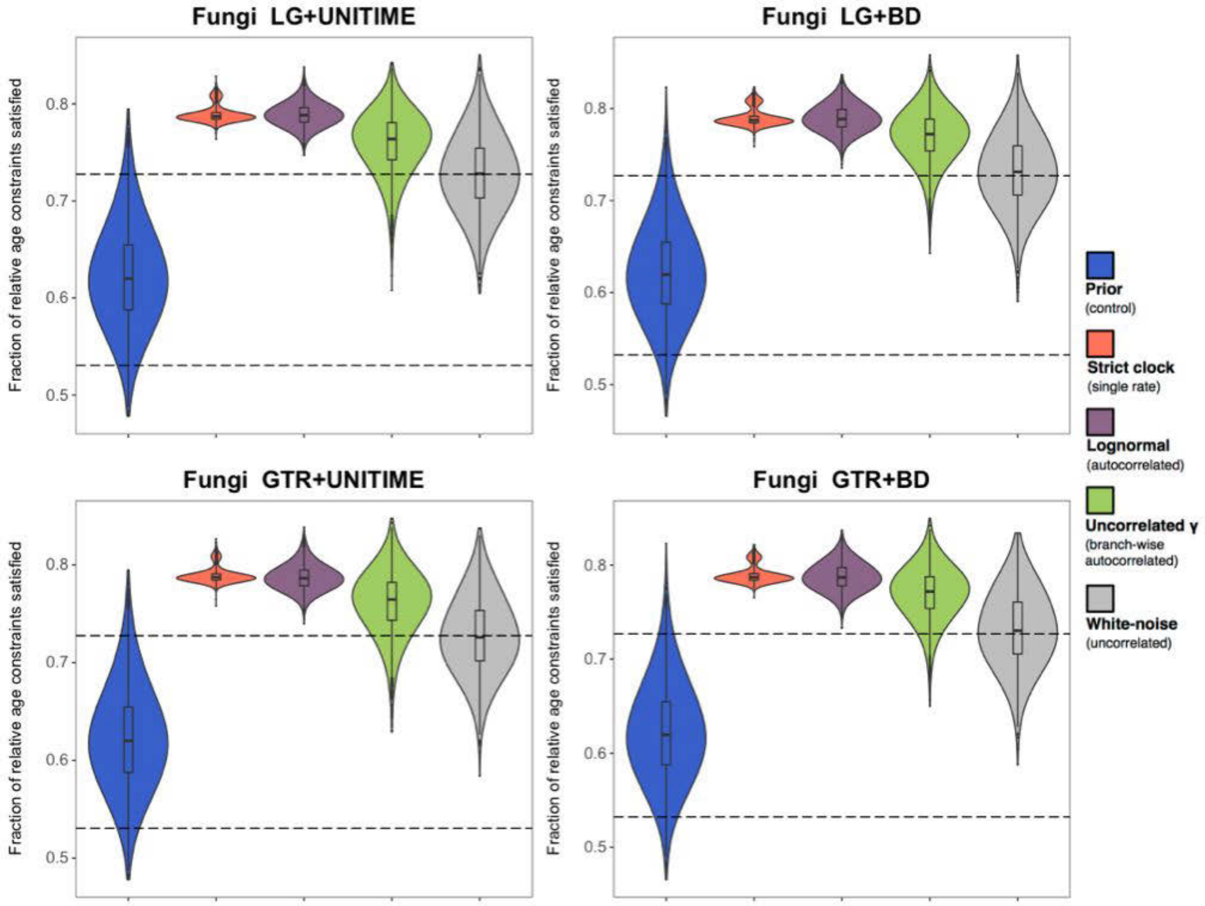

Fig. S19: Agreement between clocks and transfers based-constraints for Fungi. See text in fig S17.

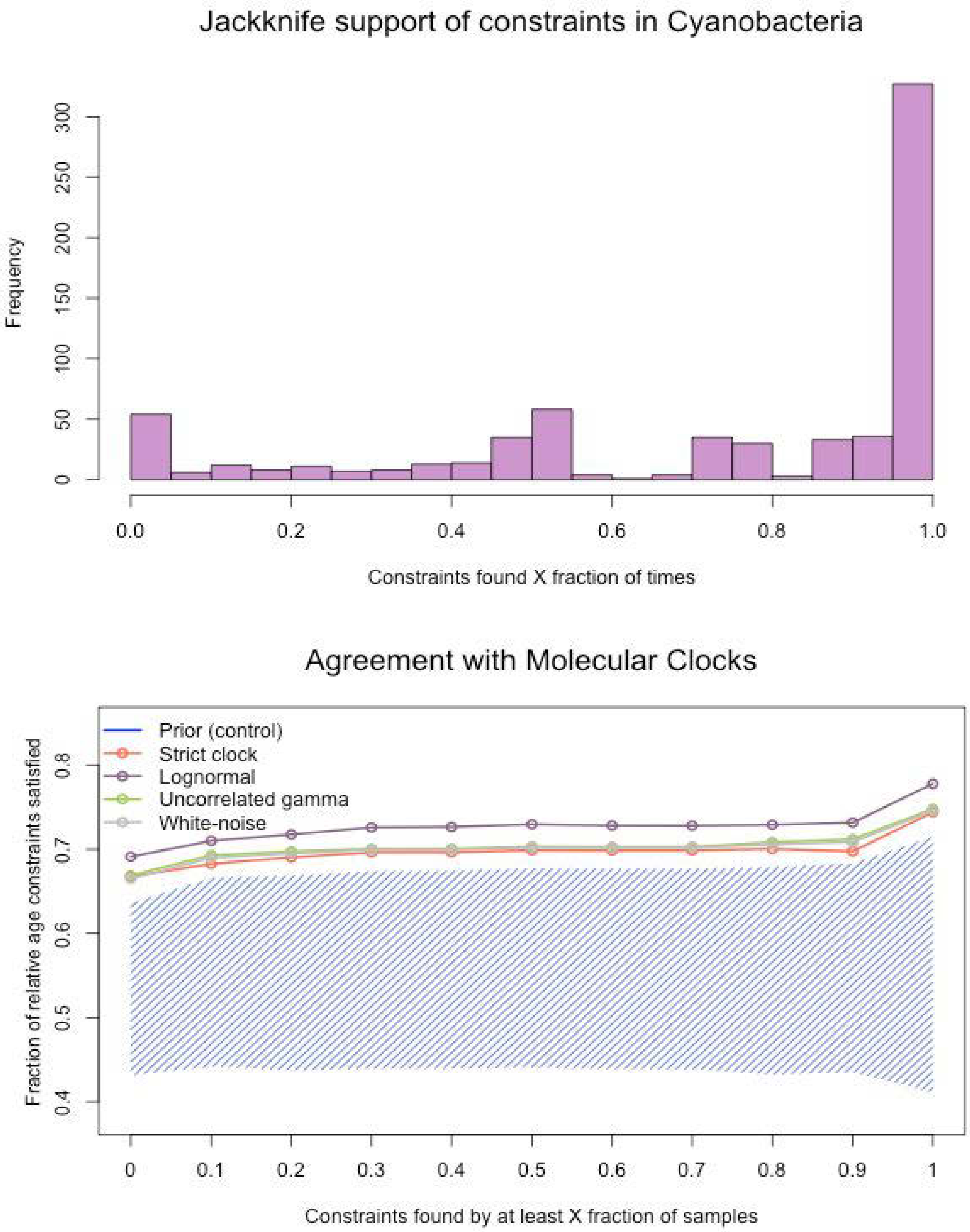

Fig. S20: Number of constraints in Cyanobacteria as a function of their support in the jackknife analysis. The constraints found at least once among all the MaxTiC trees obtained in the jackknife analysis are considered. In the lower panel we can see the agreement of those constraints with the different clock models. The blue area corresponds to the 95% area delimited by the prior, as indicated by the dashed lines in Figure 3 in the main text. The points for the different molecular clocks show the median agreement of the distribution of chronograms represented in Figure 3.

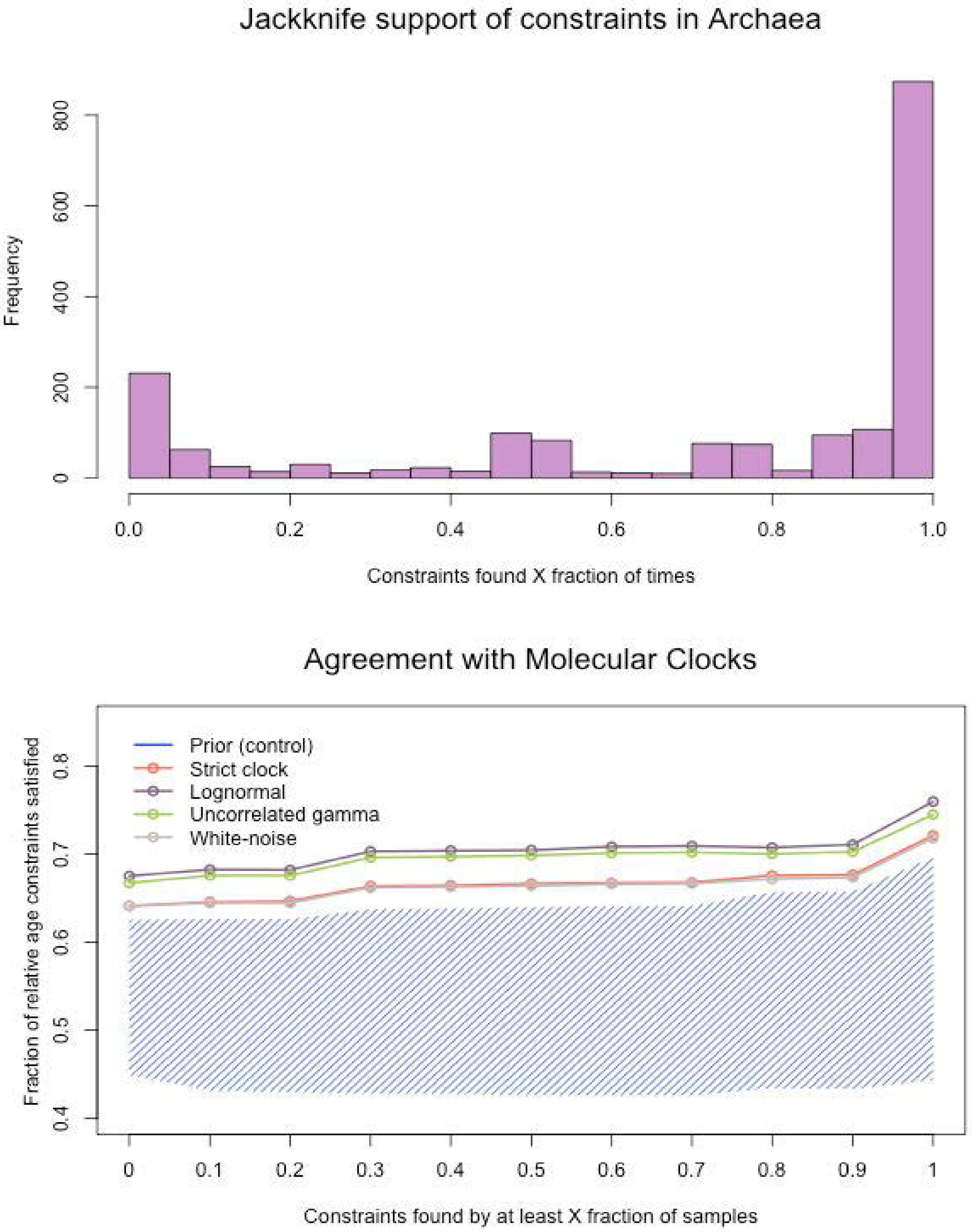

Fig. S21: Jackknife analysis for Archaea. See text in figure S20.

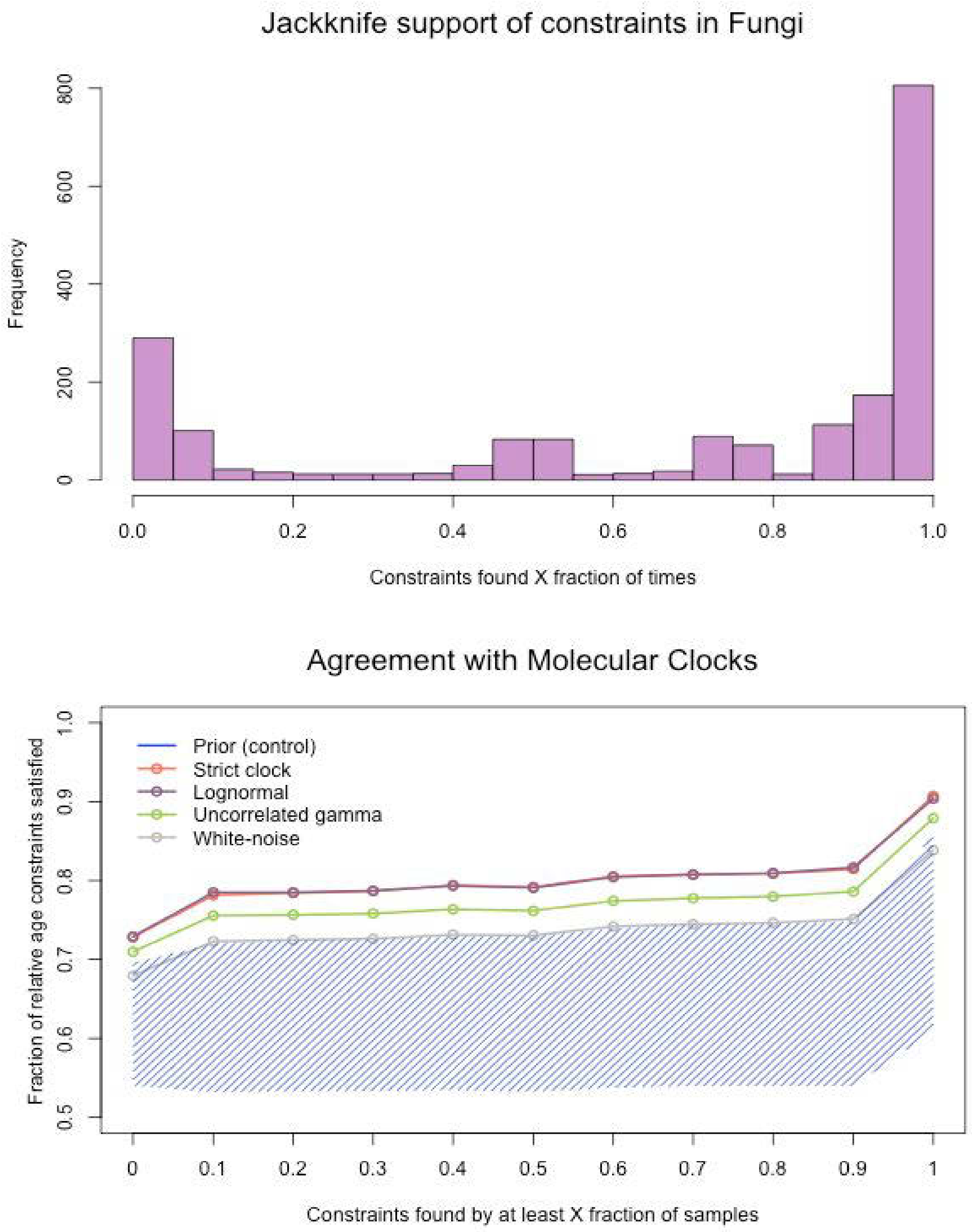

Fig. S22: Jackknife analysis for Fungi. See text in figure S20.

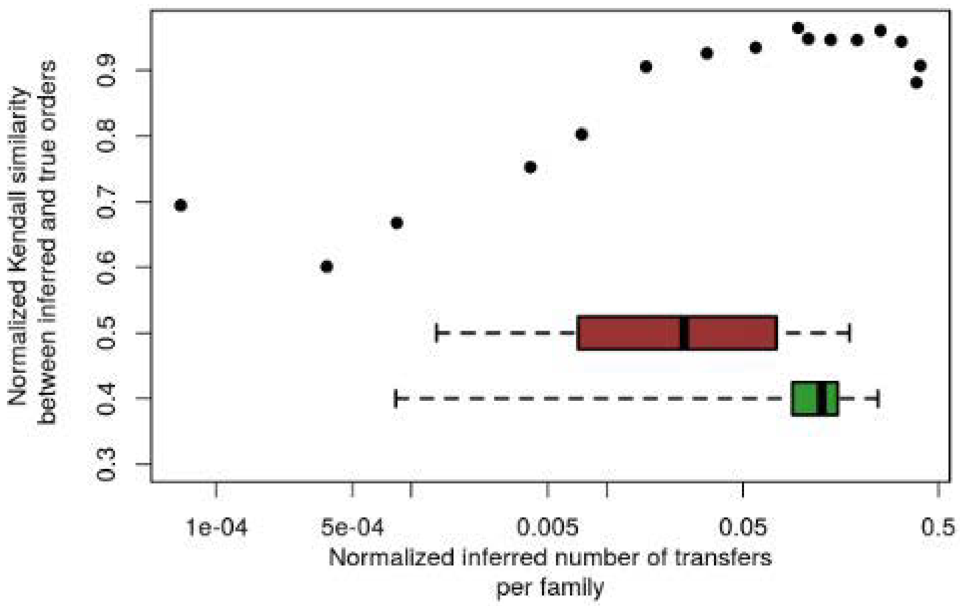

Fig. S23. Taken from Chauve et al. 2017[14]. Normalized Kendall similarity of the true ranked tree and the obtained ranked tree using MaxTiC[14], as a function of number of transfers, per branch and per family in simulated gene families (log10 scale). For each gene family the number of inferred transfers per branch is computed. In this graph we can see that for a wide range in the frequency of LGT we recover a tree very close to the real one. The levels of similarity however never reach 1, suggesting that a small fraction of transfers inferred by ALE are spurious. The boxplots correspond to the frequency of transfers found in two real datasets of Fungi and Cyanobacteria. More details on this analysis can be found in the original paper.

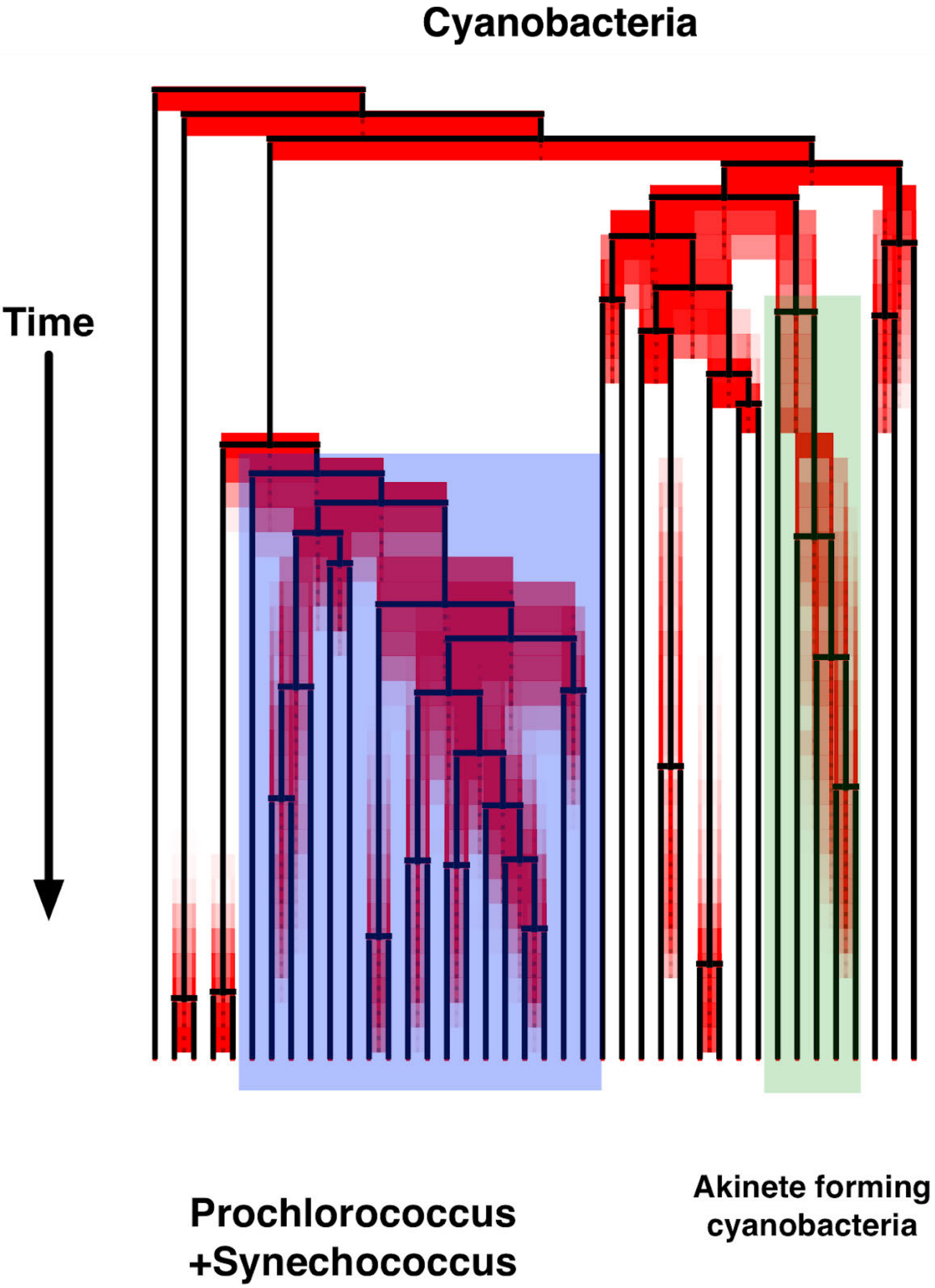

Fig. S24: Cyanobacteria species tree rooted in an alternative position, with *Gloeobacter* at the root. Most highly supported constraints (95% in a Jackknife analysis as previously described) using this tree are found also in the tree included in the main figure (183 highly supported constraints out of 191). The change of the root has a marginal effect on dating, something observed also in simulations in Chauve et al[14].

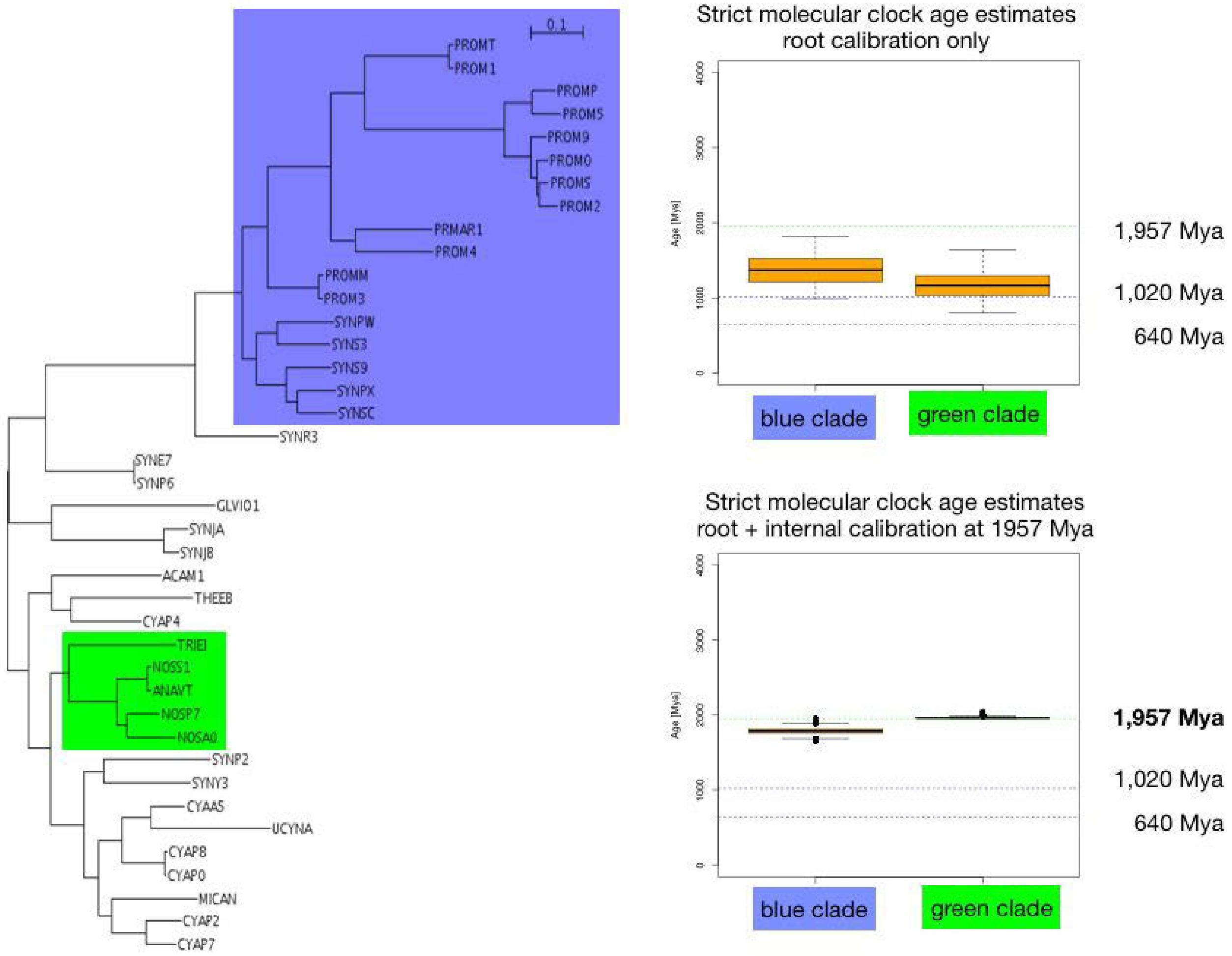

Figure. Figure S25. Phylogeny of 40 cyanobacteria with branch lengths in units of substitution per site. The blue clade is estimated to have diverged between 1,020 - 640 Mya (see e.g. Blank & Sanchez-Baracaldo, Geobiology 2010), while fossil evidence indicates that the green clade diverged at least 1,957 Mya.

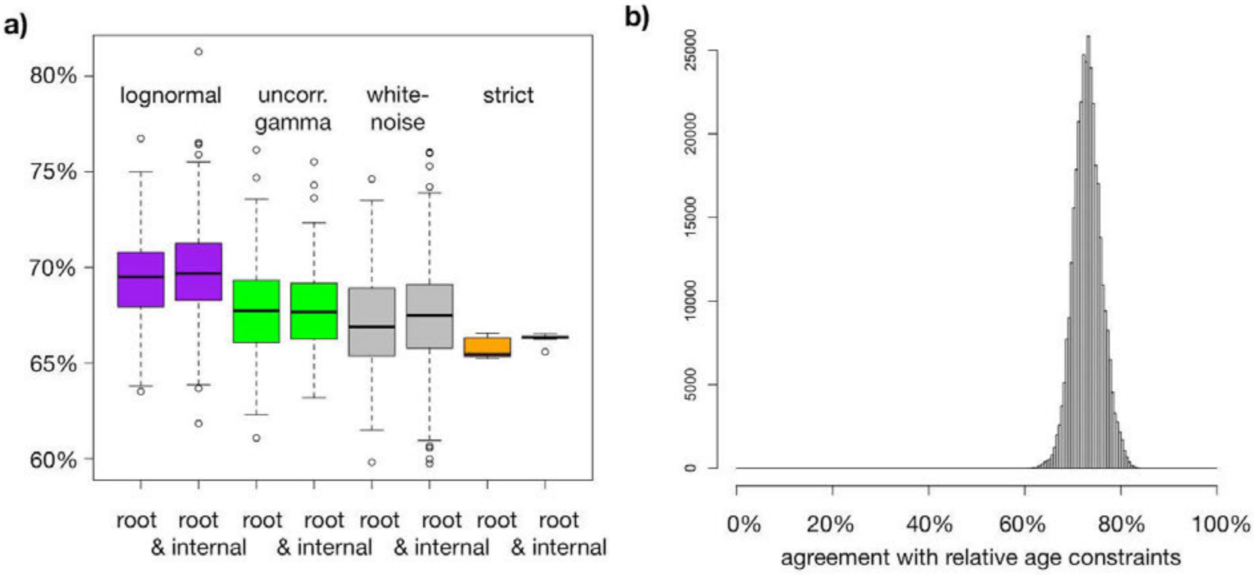

Figure S26. Introducing the internal fossil calibration increases agreement with relative transfer-based constraints. a) Agreement with transfer-based relative constraints for different relaxed molecular clock models. As before "root" indicates molecular clock models with the root constrained to be between 3,850 Mya and 2,450 Mya, while "root&internal" indicates a constraint on the age of the green clade at 1,957 Mya in addition to the root constraints. b) Agreement with transfer-based relative constraints from extensive sampling of over 300,000 chronograms under the best fitting log-normal model with "root&internal" calibrations.

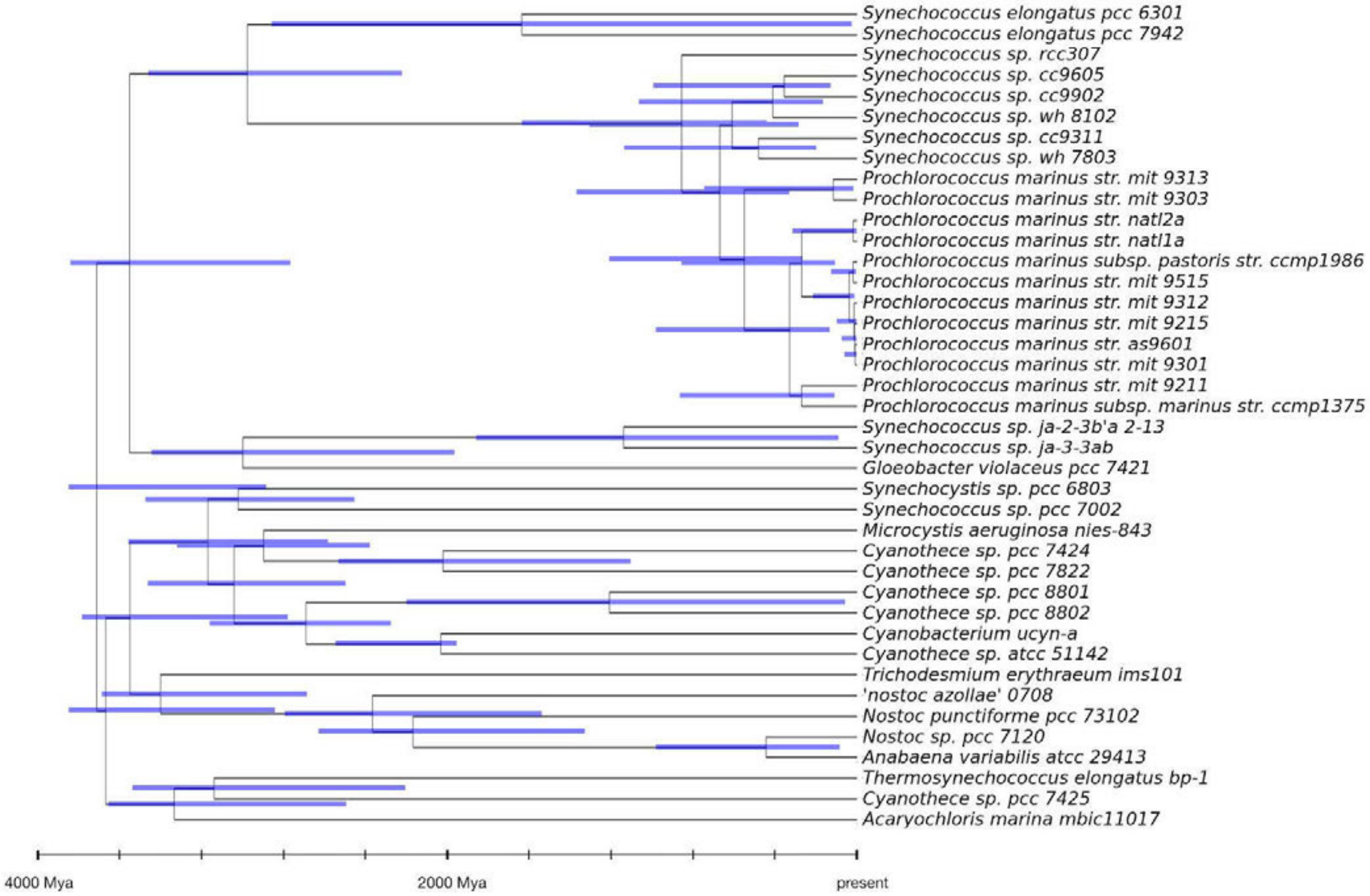
Synechococcus elongatus pcc 6301
Synechococcus elongatus pcc 7942
Synechococcus sp. rcc307
Synechococcus sp. cc9605
Synechococcus sp. cc9902
Synechococcus sp. wh 8102
Synechococcus sp. cc9311
Synechococcus sp. wh 7803
Prochlorococcus marinus str. mit 9313
Prochlorococcus marinus str. mit 9303
Prochlorococcus marinus str. natl2a
Prochlorococcus marinus str. natl1a
Prochlorococcus marinus subsp. pastoris str. ccmp1986
Prochlorococcus marinus str. mit 9515
Prochlorococcus marinus str. mit 9312
Prochlorococcus marinus str. mit 9215
Prochlorococcus marinus str. as9601
Prochlorococcus marinus str. mit 9301
Prochlorococcus marinus str. mit 9211
Prochlorococcus marinus subsp. marinus str. ccmp1375
Synechococcus sp. ja-2-3b'a 2-13
Synechococcus sp. ja-3-3ab
Gloeobacter violaceus pcc 7421
Synechocystis sp. pcc 6803
Synechococcus sp. pcc 7002
Microcystis aeruginosa nies-843
Cyanothece sp. pcc 7424
Cyanothece sp. pcc 7822
Cyanothece sp. pcc 8801
Cyanothece sp. pcc 8802
Cyanobacterium ucyn-a
Cyanothece sp. atcc 51142
Trichodesmium erythraeum ims101
'nostoc azollae' 0708
Nostoc punctiforme pcc 73102
Nostoc sp. pcc 7120
Anabaena variabilis atcc 29413
Thermosynechococcus elongatus bp-1
Cyanothece sp. pcc 7425
Acaryochloris marina mbic11017
4000 Mya
2000 Mya
present

Fig. S27: Dated species tree for Cyanobacteria based on the consensus of the subset of 5% of the sampled chronograms with the highest agreement and fossil calibrations in Table S4. The complete set of sampled chronograms and their agreement score can be downloaded from ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/

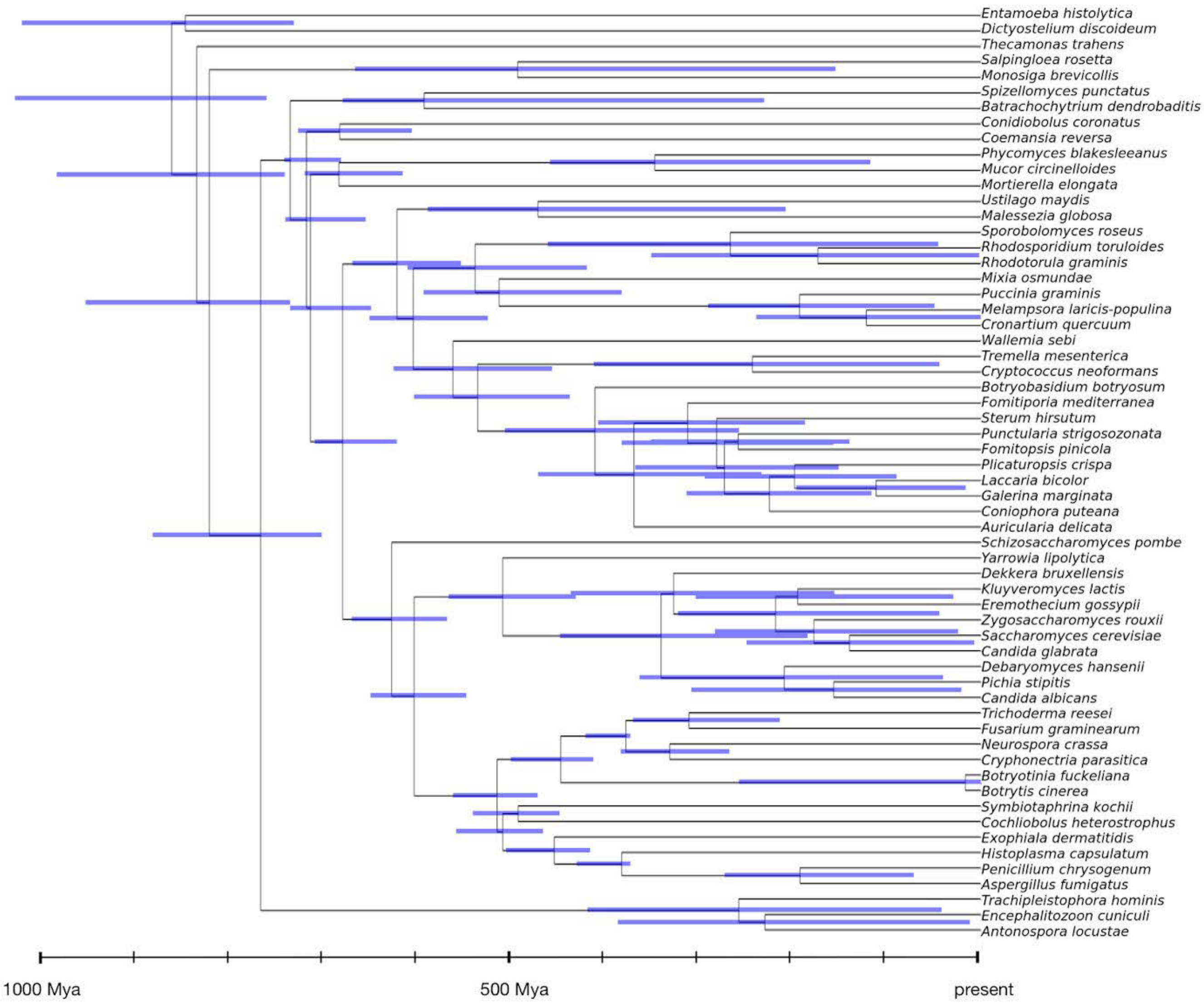

Fig. S28: Dated species tree for Fungi based on the consensus of the subset of 5% of the sampled chronograms with the highest agreement and fossil calibrations in Table S4. The complete set of sampled chronograms and their agreement score can be downloaded from ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/

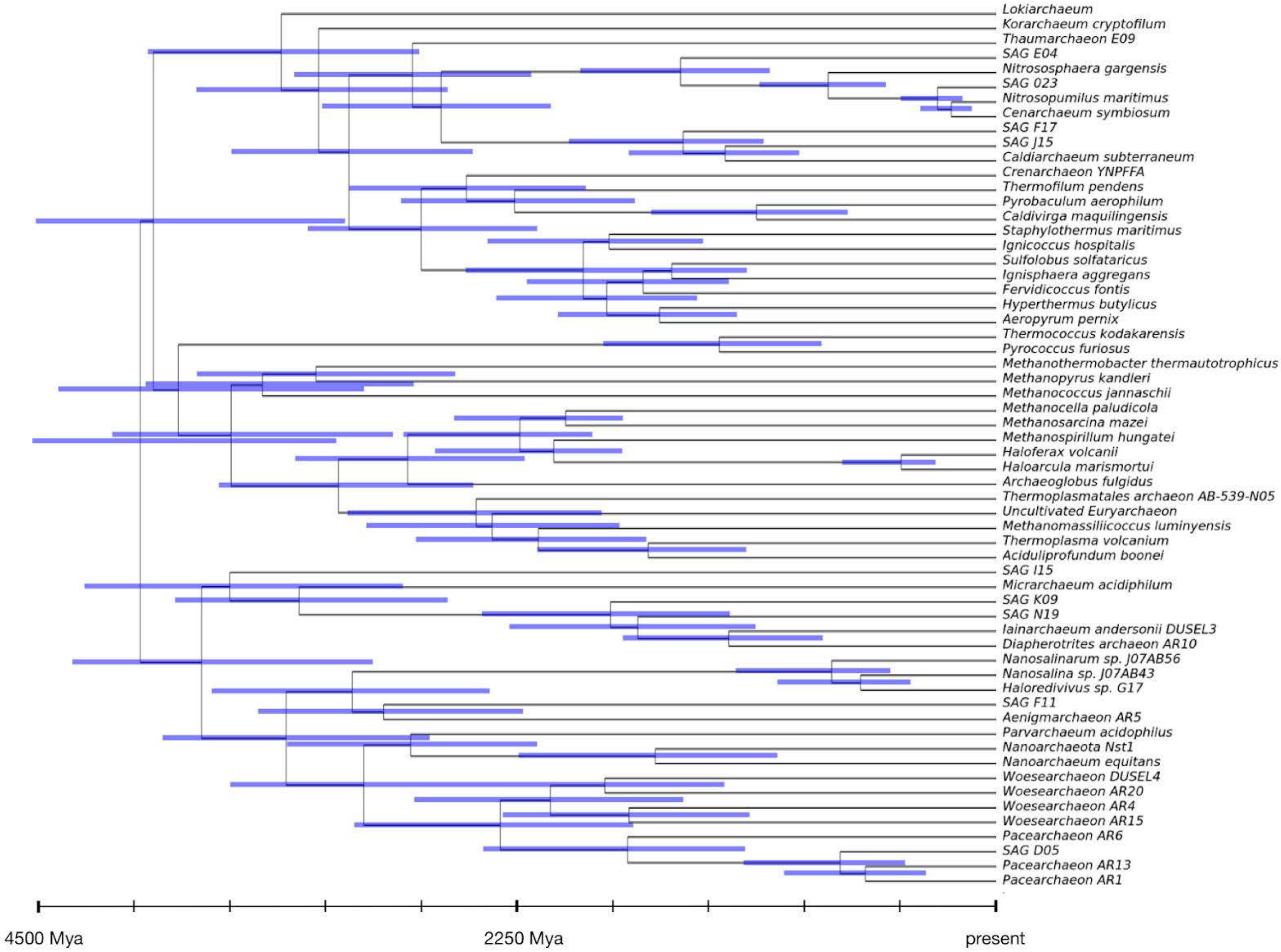

Fig. S29: Dated species tree for Archaea based on the consensus of the subset of 5% of the sampled chronograms with the highest agreement and fossil calibrations in Table S4. The complete set of sampled chronograms and their agreement score can be downloaded from ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/

Dataset	Number of genomes	Families with transfers	Total transfers (a)	Transfers filtered (b)	Constraints (c)	Conflicting constraints (d)	Size of MaxTiC (e)
Cyano	40	5322	17848.04	11084.83	4815.93	1493.06	3322.87
Archaea	60	10268	39370.58	16930.99	5123.59	1807.17	3316.42
Fungi	60	6321	35107	21800.35	7944.55	2116.77	5827.78

Table S1. Number of transfers and constraints. Each transfer is associated with a frequency that is used to weight the transfer. In a the total weight of transfers found after reconciling all families with their species tree. In b, the total weight of transfers that remain after applying the cutoff of 0.05 to remove the less supported transfers. In c the total weight of the time-informative transfers. In d, the total weight of the time informative transfers that are discarded by the MaxTiC algorithm. In e, the total weight of the time-informative transfers that are retained by the MaxTiC algorithm.

Molecular clock model	Cyanobacteria (Szöllősi) 40 genomes 3323 transfers		Archaea (Williams) 60 genomes 3316 transfers		Fungi (Nagy) 60 genomes 5828 transfers	
	A	D	A	D	A	D
Strict clock	249	79	598	276	640	166
Lognormal	261	67	646	228	638	168
Uncorrelated γ	261	67	645	229	640	166
White-noise	261	67	611	263	637	169

Table S2. Number of highly supported relative age constraints in each dataset. Constraints supported at 95% were compared with the time order of the consensus chronogram for the different molecular clocks models. They were classified as agreeing (A) if the constraint was respected by the consensus chronogram and disagreeing (D) if they were not. For a list of constraints with resampling support see Tables S5-S7.

	strict clock		white-noise		uncorr. gamma		lognormal	
calibration	root	root& internal	root	root& internal	root	root& internal	root	root& internal
blue clade < green clade	0%	100%	58%	72%	63%	95%	100%	100%
1st ancestor < green clade	0%	91%	35%	46%	44%	78%	100%	100%
2nd ancestor < green clade	0%	0%	5%	7%	3%	6%	59%	48%
3rd ancestor < green clade	0%	0%	0%	0%	0%	0%	0%	0%

Table S3. Fraction of sampled chronograms in which different transfer-based constraints are satisfied in cyanobacteria. In the calibration row, "root" indicates molecular clock models with the root constrained to be between 3,850 Mya and 2,450 Mya, while "root&internal" indicates a constraint on the age of the green clade at 1,957 Mya in addition to the root constraints.

Dataset	Node	Event	Minimum age (mya)	Maximum age (mya)	Reference
Cyanobacteria	67	*Akinete forming cyanobacteria*	1957	3850	26
	Root	*Great Oxygenation Event*	2450	3850	27
Archaea	116	*Oldest Eukaryotes*	1776	-	28
Fungi	111	*Dikarya*	392.1	-	29
	87	*Paleopyrenomycites*	400	-	30
	114	*Divergence of Chytridiomycota*	-	750	31
	Root	*Oldest Eukaryotes*	-	1891	28

Table S4. Fossils calibrations used.

Attached as separate files:
Table S5. Constraints and their support in Cyanobacteria
Table S6. Constraints and their support in Archaea
Table S7. Constraints and their support in Fungi

Additional material can be downloaded from:
ftp://pbil.univ-lyon1.fr/pub/datasets/davin2017/

References

1. dos Reis, M. *et al.* Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society B: Biological Sciences* **279,** 3491–3500 (2012).
2. Szöllősi, G. J., Davín, A. A., Tannier, E., Daubin, V. & Boussau, B. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370,** 20140335 (2015).
3. Williams, T. A. *et al.* Integrative modelling of gene and genome evolution roots the archaeal tree of life. *PNAS* (2017).
4. Nagy, L. G. *et al.* Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5,** 4471 (2014).
5. Larget, B. The Estimation of Tree Posterior Probabilities Using Conditional Clade Probability Distributions. *Syst. Biol.* **62,** 501–511 (2013).
6. Höhna, S. & Drummond, A. J. Guided tree topology proposals for Bayesian phylogenetic inference. *Syst. Biol.* **61,** 1–11 (2012).
7. Szöllősi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E. & Daubin, V. Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62,** 901–912 (2013).
8. Szöllősi, G. J., Boussau, B., Abby, S. S., Tannier, E. & Daubin, V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *PNAS* **109,** 17513–17518 (2012).
9. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25,** 2286–2288 (2009).
10. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25,** 1307–1320 (2008).
11. Szöllősi, G. J., Szoll si, G. J., Tannier, E., Daubin, V. & Boussau, B. The Inference of Gene

Trees with Species Trees. *Syst. Biol.* **64,** e42–e62 (2014).

12. Szöllősi, G. J., Tannier, E., Lartillot, N. & Daubin, V. Lateral Gene Transfer from the Dead. *Syst. Biol.* **62,** 386–397 (2013).
13. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33,** 1635–1638 (2016).
14. Chauve, C. *et al.* MaxTiC: Fast ranking of a phylogenetic tree by Maximum Time Consistency with lateral gene Transfers. *bioRxiv* (2017). doi: https://doi.org/10.1101/127548
15. Kishino, H., Thorne, J. L. & Bruno, W. J. Performance of a Divergence Time Estimation Method under a Probabilistic Model of Rate Evolution. *Mol. Biol. Evol.* **18,** 352–361 (2001).
16. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* **4,** e88 (2006).
17. Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. A General Comparison of Relaxed Molecular Clock Models. *Mol. Biol. Evol.* **24,** 2669–2680 (2007).
18. Lanave, C., Preparata, G., Saccone, C. & Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20,** 86–93 (1984).
19. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32,** 268–274 (2015).
20. Abby, S. S., Tannier, E., Gouy, M. & Daubin, V. Lateral gene transfer as a support for the tree of life. *Proceedings of the National Academy of Sciences* **109,** 4962–4967 (2012).
21. Swingley, W. D., Blankenship, R. E. & Raymond, J. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol. Biol. Evol.* **25,** 643–654 (2008).
22. Criscuolo, A. & Gribaldo, S. Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol. Biol. Evol.* **28,** 3019–3032 (2011).
23. Schirrmeister, B. E., Antonelli, A. & Bagheri, H. C. The origin of multicellularity in

cyanobacteria. *BMC Evol. Biol.* **11,** (2011).

24. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nature Ecology & Evolution* **1,** s41559–017 (2017).

25. Jacox, E., Chauve, C., Szöllősi, G. J., Ponty, Y. & Scornavacca, C. ecceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinformatics* **32,** 2056–2058 (2016).

26. Tomitani, A., Knoll, A. H., Cavanaugh, C. M. & Ohno, T. The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proceedings of the National Academy of Sciences* **103,** 5442–5447 (2006).

27. Blank, C. E. & Sanchez-Baracaldo, P. Timing of morphological and ecological innovations in the cyanobacteria â a key to understanding the rise in atmospheric oxygen. *Geobiology* **8,** 1–23 (2010).

28. Kusky, T. M. & Li, J. Paleoproterozoic tectonic evolution of the North China Craton. *J. Asian Earth Sci.* **22,** 383–397 (2003).

29. Gradstein, F. M., Ogg, J. G., Schmitz, M. & Ogg, G. *The Geologic Time Scale 2012*. (Elsevier, 2012).

30. Taylor, T. N., Hass, H., Kerp, H., Krings, M. & Hanlin, R. T. Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. *Mycologia* **97,** 269–285 (2005).

31. Chang, Y. *et al.* Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biol. Evol.* **7,** 1590–1601 (2015).